Studies of stored mRNAs during seed aging in *Arabidopsis thaliana*, *Brassica napus* and wheat (*Triticum aestivum*)

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Biochemistry, Microbiology and Immunology

University of Saskatchewan

Saskatoon

By

Liang Zhao

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis. Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

OR

Head
Department of Biochemistry, Microbiology and Immunology
University of Saskatchewan
107 Wiggins Road
Saskatoon, Saskatchewan, S7N 5E5
Canada

ABSTRACT

How plant seeds age remains poorly understood and effective tools for monitoring seed aging are lacking. Dry seeds contain various stored mRNAs which are believed to be required for protein synthesis during early stages of seed germination. We reasoned that seed stored mRNAs would undergo degradation during seed aging, based on previous research and the propensity of mRNAs to degrade. I performed RT-PCR, qPCR, and MinION nanopore sequencing analyses to study the changes in stored mRNAs of Arabidopsis seeds, and the hypothesis was approved by the following results. First, all stored mRNAs analyzed were gradually degraded in naturally and acceleratedly aged seeds. The difference in the threshold cycle (Ct) number of qPCR analysis between aged and control seeds (ΔCt value) was highly correlated with the mRNA fragment size and seed aging time. Second, mathematical equations were derived for estimating the relative amount of undamaged stored mRNAs and frequency of the breakdown at the one nucleotide level for individual mRNAs. Third, results indicated that stored mRNAs were broken down randomly. The frequency of breaks per nucleotide per day, named as a β value, remained fairly constant over the aging time under the same aging conditions, but increased greatly with temperature. Fourth, RNA-seq analysis using MinION nanopore sequencing revealed a genome-wide trend of decreasing stored mRNA lengths in aged seeds. However, the changes were not as profound as observed with the qPCR analysis. Fifth, the method based the ΔCt value reflecting stored mRNA change was found to be more precise than three existing methods for seed aging assessment. The findings in Arabidopsis were extended with similar observations in wheat and canola seeds. The ΔCt value also highly correlated with the mRNA fragment size and seed aging time for these two crop seeds. Interestingly, the β values for the three plant species were generally similar suggesting comparable rates of stored mRNA degradation under the same conditions. These observations have raised interesting questions on the traditionally proposed mechanisms for seed aging and differences in seed aging among plants. Further, the methods developed here should be useful for studying stored mRNAs and seed aging.

ACKNOWLEDGEMENTS

First of all, I would like to give my thanks to my supervisors Dr. Hong Wang and Dr. Yong-Bi Fu (Plant Gene Resources of Canada, Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada), for their continuous support and encouragement during my study. Dr. Wang and Dr. Fu gave me thorough guidance on my scientific research and strong support when I met problems. Especially, I would like to give my special thanks to Dr. Wang for the great help he provided for me to finish the program and the efforts he made in revising my thesis

I would like to thank my advisory committee members Dr. Jeremy Lee (the chair), Dr. Stanley Moore, Dr. Patrick Covello, Dr. Bill Roesler, Dr. Gopalan Selvaraj, and Dr. Karen Tanino for their valuable feedback and suggestions. Especially, I would like to express my gratitude to Dr. Tanino for her time and efforts in improving my English and preparing for my comprehensive exam, even when she was very busy at the time.

Moreover, I appreciate the help and friendship of my lab mates: Dr. Sheng Wang for the initial RNA isolation and cDNA synthesis; Dr. Ling Cao for taking care of plants when I was away; and Shengjian Ye for general lab help.

I would also like to thank Dr. Yong-Bi Fu and Dr. Hong Wang for the valuable financial support to me and for my research, and the Department of Biochemistry for providing a devolved scholarship for three years.

Finally, my special thanks are given to my family, my father Peicheng Zhao, my mother Anqiong Zheng, my brother Fei Zhao, and my sister Lijuan Zhao. Their unconditional love and encouragement are my source of motivation.

TABLE OF CONTENTS

PERMISSION TO USE i

ABSTRACT ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF ABBREVATIONS xii

1. CHAPTER ONE - INTRODUCTION AND LITERATURE REVIEW 1

1.1 Introduction 1

1.2 Main factors affecting seed aging and longevity 4

1.2.1 Seed glassy state 4

1.2.2 Reactive oxygen species 6

1.2.3 Damage to lipids and membrane 8

1.2.4 Mitochondrial damage 10

1.2.5 Damage to protein 11

1.2.6 Damage to DNA 13

1.2.7 Stored mRNAs and seed aging 14

1.2.7.1 Seed stored mRNAs 14

1.2.7.2 Roles of seed stored mRNAs 15

1.2.7.3 Damage to seed stored mRNAs 17

1.2.8 Pathways of mRNA degradation 18

1.2.9 Protective compounds or biological processes 20

1.2.10 Seed coat 21

1.2.11 Typical methods in monitoring seed aging 22
1.3 Research questions, hypothesis, and objectives 22
2 . MATERIALS AND METHODS 25
2.1 Seed materials 25
2.2 Seed aging treatments 25
2.3 Isolation of total RNA from seeds 27
2.4 Analysis of total RNA integrity 28
2.5 Reverse transcription PCR (RT-PCR) and quantitative PCR (qPCR) analyses 28
2.6 Estimating the relative amount of undamaged stored mRNAs during seed aging 33
2.7 Estimating the rate of stored mRNA degradation at the one nucleotide level 33
2.8 Electrical conductivity (EC) Test 35
2.9 Analyses of seedling growth 35
2.10 MinION nanopore sequencing 36
2.10.1 cDNA synthesis, template enrichment, and purification 36
2.10.2 cDNA end-repair 37
2.10.3 Adapter ligation to cDNA ends 37
2.10.4 Preparing preloading mixture and priming flow cell 37
2.10.5 Acquisition, processing and analysis of the sequencing reads 38
3. CHAPTER THREE – SEED STORED mRNA DEGRADATION IN ARABIDOPSIS SEEDS 39
3.1 Introduction 39
3.2 Results 40
3.2.1 Seed aging treatments 40
3.2.2 Comparison of methods for RNA isolation from Arabidopsis seeds 42

3.2.3 Improvement of an RNA isolation method 42
3.2.4 RT-PCR analysis of seed stored mRNA levels in aged Arabidopsis seeds 46
3.2.5 qPCR analysis of seed stored mRNA levels in aged seeds 52
3.2.6 Analysis of the relationship between ΔCt value and mRNA fragment length in aged seeds 55
3.2.7 Analysis of the relationship between the decrease in stored mRNAs and seed germination 56
3.2.8 Development of a quantitative method for determining the changes in stored mRNA levels during seed aging 59
3.2.9 Estimating the relative amount of undamaged stored mRNAs during seed aging 62
3.2.10 Half-life time for different fragments 64
3.2.11 Estimating the rate of stored mRNA degradation at the one nucleotide level 65
3.2.12 MinION nanopore sequencing of cDNA derived from seed stored mRNAs 69
3.2.13 Comparisons with traditional methods of assessing seed aging 73
3.3 Discussion 75
3.3.1 Degradation of stored mRNAs is highly correlated with aging in naturally and artificially aged Arabidopsis seeds 75
3.3.2 A new approach of normalizing qPCR data improves the analysis of stored mRNAs 77
3.3.3 Quantitative methods are developed to estimate the relative amount of intact mRNAs and rate of mRNA degradation at the one nucleotide level 77
3.3.4 MinION sequencing results also show the degradation of stored mRNAs in seed aging 79
3.3.5 Stored mRNAs can serve as more precise biomarkers for monitoring seed aging 80
4. CHAPTER FOUR - SEED STORED mRNA DEGRADATION IN WHEAT AND

CANOLA SEEDS 82
4.1 Introduction 82
4.2 Seed stored mRNA degradation in wheat seeds 82
4.2.1 Comparison of wheat germination assays 82
4.2.2 Wheat accelerated aging assay 83
4.2.3 RT-PCR analysis of stored mRNAs during wheat seed aging 85
4.2.4 qPCR analysis of stored mRNAs in aged wheat seeds 86
4.3 Seed stored mRNA degradation in *B. napus* seeds 90
4.3.1 Decreases of seed viability with aging time for *B. napus* 90
4.3.2 Survey of the presence of stored mRNAs in *B. napus* seeds 91
4.3.3 Quantitative analysis of seed stored mRNA in aged *B. napus* seeds 92
4.4 Discussion 95
4.4.1 New methods are useful for quantifying stored mRNA degradation 95
4.4.2 Comparisons of stored mRNA degradation and seed aging in three different plant species raise some interesting questions 97
5. CHAPTER FIVE - GENERAL DISCUSSION AND FUTURE WORK 100
5.1 General discussion 100
5.1.1 Importance and current status of seed aging research 100
5.1.2 Novel findings have been made and methods developed on stored mRNA degradation during seed aging in three different plants 101
5.1.3 Current findings challenge some traditional concepts on seed aging 103
5.1.2 Practical implications for germplasm management 104
5.2 Future work 106
5.2.1 Extending the studies to more plant species 106

5.2.2 Stored mRNAs and seed aging 107

SUPPLEMENTARY FILE 123

Supplementary file 1: Perl scripts used in MinION nanopore sequencing 123

LIST OF TABLES

Table 2.1 List of naturally aged Arabidopsis seeds used in this study. 26

Table 2.2 List of genes or fragments used in qPCR analysis on changes in stored mRNAs in aged Arabidopsis seeds..... 30

Table 2.3 List of genes (or fragments) and primers used to analyze mRNA degradation in wheat seeds. 31

Table 2.4 List of *B. napus* genes (or fragments) and primers used to study stored mRNA degradation during seed aging. 32

Table 3.1 RNA quality test for the total RNAs isolated from three methods 43

Table 3.2 RNA quality test for the total RNAs isolated from unaged, AA and NA seeds using the modified method. 45

Table 3.3 RNA integrity numbers (RINs) of RNAs isolated from unaged and aged Arabidopsis seeds..... 46

Table 3.4 List of genes used to survey the presence of seed stored mRNAs by RT-PCR. 47

Table 3.5 ΔCt values for 29 genes in qPCR analysis on changes in stored mRNAs in aged Arabidopsis seeds..... 43

Table 3.6 Estimated β values for different fragments at different aging times..... 67

Table 3.7 Estimated β values for stored mRNAs of six genes at different aging times..... 67

Table 3.8 Estimated fragment β values for Arabidopsis seeds aged at different aging temperatures..... 68

Table 4.1 Estimated β values for different wheat fragments at different aging times. 89

Table 4.2 Estimated β values for different *B. napus* fragments at different aging times. 95

LIST OF FIGURES

Figure 3.1 Effect of accelerated aging conditions on germination percentage of freshly harvested Arabidopsis seeds. 41

Figure 3.2 Effect of natural storage on seed germination percentage. 41

Figure 3.3 Comparison of three protocols for RNA isolation from dry Arabidopsis seeds. ... 43

Figure 3.4 Example of total RNAs isolated from unaged and aged Arabidopsis dry seeds.... 44

Figure 3.5 Results of RT-PCR analysis of 120 genes (in A, B, C and D groups) to survey the presence of stored mRNAs in Arabidopsis dry seeds. 48

Figure 3.6 The level of stored mRNAs in naturally aged (NA) and acceleratedly aged (AA) Arabidopsis seeds detected by RT-PCR. 51

Figure 3.7 The relationship between the length of Arabidopsis stored mRNAs and ΔCt value in qPCR analysis. 54

Figure 3.8 Diagram to show the positions of different fragments on the cDNA derived from a stored mRNA. 55

Figure 3.9 The relationship between ΔCt value and cDNA fragment size analyzed using the same genes. 57

Figure 3.10 Relationship between ΔCt for the cDNA (2500 bp fragment) of gene B16 and Arabidopsis seed germination percentage. 58

Figure 3.11 Change in the percentage of germination for Arabidopsis seeds following the accelerated aging treatment. 59

Figure 3.12 Correlation of mRNA degradation with seed aging time analyzed using different cDNA fragment lengths of two genes. 61

Figure 3.13 The relationship between the ΔCt values from analyzing stored mRNAs of six genes and seed aging time. 63

Figure 3.14 The estimated percentage of undamaged mRNAs of B16 and B20 in seeds with different accelerated aging time (days). 65

Figure 3.15 The estimated half-life times of different fragments from $B16_{2250bp}$ and $B20_{1750bp}$. 65

Figure 3.16 Estimated β values for stored mRNAs in seeds aged at different temperatures. . 69

Figure 3.17 Comparison of read coverage between unaged and AA sample. ……72

Figure 3.18 Analysis of seed aging by traditional methods. 74

Figure 4.1 Comparison of two wheat seed germination assays. 84

Figure 4.2 Change in germination percentage for wheat seeds aged in an accelerated aging assay. ……...85

Figure 4.3 Results of RT-PCR analysis of 19 genes to survey the presence of stored mRNAs in wheat dry seeds. 86

Figure 4.4 The relationship between ΔCt value and cDNA fragment size analyzed using one wheat gene. 87

Figure 4.5 Correlation of mRNA degradation with seed aging time analyzed using different cDNA fragment lengths of gene W2. 88

Figure 4.6 The relationship between the ΔCt values of stored mRNAs of four wheat genes and seed aging time. 90

Figure 4.7 The change in germination percentage in *B. napus* seeds with AA days. 91

Figure 4.8 The presence of stored mRNAs in *B. napus* dry seeds for the genes listed in Table 2.4. 91

Figure 4.9 The relationship between ΔCt value and cDNA fragment size analyzed using *B. napus* gene Bn12. 92

Figure 4.10 Correlation of mRNA degradation with seed aging time analyzed using different cDNA fragments of Bn12. 93

Figure 4.11 The relationship between the ΔCt values of stored mRNAs of six *B. napus* genes and seed aging time. 94

LIST OF ABBREVATIONS

8-oxo-G	7, 8-dihydro-8-oxoguanine
AA	accelerated aging
ActD	actinomycin D
AGO	Argonaute
ANOVA	analysis of variance
APX	ascorbate peroxidase
AtCAF1A	Arabidopsis CCR4-associate-factor 1A
CCR4-NOT	carbon catabolite repressor 4-negative on TATA complex
Ct	threshold cycle
CuZnSOD	Cu/Zn-superoxide dismutase
DAF	days after flowering
DCP1	Decapping Protein 1
dsDNA	double-stranded dsDNA
ddPCR	droplet digital PCR
EC	electrical conductivity
EJC	exon-exon junction complex
ELF3	Early Flowering 3
FAO	Food and Agriculture Organization
FLM	Flowering Locus M
GSH	glutathione
GSH-Px	glutathione peroxidase
HSPs	heat shock proteins
LEA	late embryogenesis abundant
LOX2	lipoxygenase 2
LOX3	lipoxygenase 3
M1G	pyrimidopurinone
MS	Murashige and Skoog

NA	natural aging
NADPH	nicotinamide adenine dinucleotide phosphate
NGS	next generation sequencing
NGD	non-go decay
NMD	nonsense-mediated decay
NSD	non-stop decay
OGG1	8-oxoguanine-DNA glycosylase 1
PAN2	poly(A) ribonuclease 2
PCD	programmed cell death
PCR	polymerase chain reaction
PIMT	protein-isoaspartyl-methytransferase
PLD	phospholipase D
PTC	premature stop codon
PUFAs	polyunsaturated fatty acids
qPCR	real-time quantitative PCR
RCCs	reactive carbonyl compounds
RISC	RNA-induced silencing complex
ROS	reactive oxygen species
RT-PCR	reverse transcription PCR
RTase	reverse transcriptase
SOD	superoxide dismutase
ZMW	zero-mode waveguide
ΔCt	delta Ct

1. CHAPTER ONE - INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Seed longevity is defined as the total life span of seeds retaining the ability to germinate (Nguyen *et al.*, 2012). Seed longevity varies greatly among different species and within a species depending on the storage conditions. The huge variance is likely the consequence of natural selection by the environmental conditions in which plants live. Some physiological and morphological features exhibited by plant seeds may partly reflect the way they adapt to the natural conditions or the way their ancestors survived the past environmental conditions (Vázquez-Yanes and Orozco-Segovia, 1993).

Seeds longevity is obviously important for the succession of a plant species, although it may not be equally important for different plants. The succession of a plant species depends on its abilities to reproduce, disseminate in the same environmental territories, move or be carried to new environment, and survive unfavorable environmental conditions. The relative importance of these abilities depends on the environmental conditions which impose constraints on different plant species. For example, both high temperature and high moisture in the rainforest select for seeds with strong abilities to reproduce and disperse (Vázquez-Yanes and Orozco-Segovia, 1993). The ability to survive long periods of conditions unfavourable for seed germination becomes less important compared to mesotherm plant seeds which need to survive harsh winter conditions. Thus, seeds in rainforest generally exhibit shorter lifespan than their counterparts in the mesotherm (Vázquez-Yanes and Orozco-Segovia, 1993).

Seed longevity is also critical for germplasm conservation. Germplasm resources provides much needed genetic resources in exploring the uses of plants for food, fibre and feed, but plant genetics resources are increasingly being lost [Food and Agriculture Organization (FAO), 2010]. Recently, Humphreys *et al*., (2019) reported that seed plants extinct at a faster rate after 1900 AD than history records before 1900. In an effort to conserve genetic resources, plant germplasms could be maintained through *in situ* (e.g. protected areas, natural parks) or *ex situ* (e.g. tissue culture collections, seed banks, and

botanic gardens) approaches (Li and Pritchard, 2009). For plant seeds, *ex situ* conservation in seed banks is economically more preferable. Currently, about 7.4 million seed accessions are collected and protected in various seedbanks and botanical gardens (FAO, 2010). However, seed vigor and viability decrease with storage time and the maximum storage duration (or seed longevity) even under the low temperature genebank conditions (usually −18 °C) is not unlimiteded (Ventura *et al*., 2012). Seeds vigor refers to a complex trait of rapid and uniform seedling emergence in various field conditions, and it depends on the ability to resist long-term storage and seed aging (Ventura *et al*., 2012). Seed samples in seed banks need to be regenerated when the viability declines to a certain level, which is usually evaluated based on seed germination assays and set to be 65% to 85% of seed germination, depending on the specific seed banks (Walters, 2015). Thus, the seed samples need to be frequently tested to ensure that they are regenerated before their ability to germinate is lost completely, which is a challenging task considering the huge number of seed samples in the seed banks (Walters, 2015).

Seed aging could result in reduced seed vigor and affects agricultural production. Seeds start to age shortly after their physiological maturity in the field and the aging process continues during storage (Ellis *et al*., 1987; Ellis and Roberts, 1981). Before seeds lose the viability completely, changes occurring during seed aging could result in slower germination and sub-optimal performances in seedling establishment (Ellis and Roberts, 1981; Khah *et al*., 1989; Shaban, 2013), which are generally referred to as reduced or poor seed vigor (Perry, 1984). The poor seed vigor could reduce agricultural production by reducing the percentage of seed germination and plant density in the field. Even if the germination percentage and plant density are unaffected, the reduced seed vigor could result in delayed seedling emergence and smaller seedlings (Gray, 1976; Salter *et al*., 2009; Wurr, 1983) which are disadvantageous in competing for nutrition and light (Tseng and Lin, 1962; Wester, 1966). Furthermore, the aged seeds may have more genetic aberrations which lead to impaired physiological functions (TeKrony and Egli, 1991). The above mentioned negative effects caused by seed aging might lead to a reduction in crop yield depending on the crop types (David, 1987). For crops harvested during vegetative growth (e.g. sugarbeet and cabbage)

and at an early stage of reproductive development (e.g. pea and tomato), the uneven germination and delayed seedling emergence may result in reduced plant uniformity and yield (TeKrony and Egli, 1991).

However, seed aging is not always detrimental to seeds. First, short term storage or moderate seed aging could promote seed germination through breaking dormancy and loosening seed coat and endosperm. Seed dormancy refers to a temporary failure of seed germination under favorable conditions, although the seeds are still viable (Bewley, 1997). Seed dormancy is a wide-spread phenomenon in seed plants with the extent varying greatly among different species (Nguyen *et al.*, 2012), and it is an evolutionarily adapted trait that helps to delay and stagger seed germination to increase the survival of a particular species. For many plant species, dormancy can be removed (or mitigated) by dry storage (referred to as after-ripening) or seed stratification (e.g. storing the water-saturated Arabidopsis seeds in a 4-5 °C refrigerator for 2 days in this study). Second, storage materials in seeds such as carbohydrates and proteins may be slightly oxidized during seed aging, which could facilitate their mobilization during seed germination. Thus, moderate seed aging could be helpful in uniform germination by breaking seed dormancy. The focus of this research project however is on seed aging.

With prolonged storage and seed aging, seeds eventually lose the viability. Seed viability refers to the capability of seeds to germinate under suitable conditions, which also includes dormant but viable seeds. Since after-ripening and seed stratification are conducted usually prior to seed germination tests, seed germination is the most common method to evaluate seed viability and seed aging status. Although seed germination assay is easy to perform, seed germination is based on a singular break-point event that a seed no longer is able to germinate and it reveals little regarding what happens during seed storage until this point. In this regard, Fu e*t al*. (2017) showed that the traditional germination tests were not so accurate in assaying more than 500 accessions of wheat seeds. Further, the germination tests usually are conducted under laboratory conditions which are different from field conditions; and the results may not fully reflect seed viability in the field (Corbineau, 2012). Other methods of assessing seed aging have been developed such as seedling growth,

tetrazolium test, electrical conductivity test, and ROS scavenging ability (Fu *et al*., 2015). However, most of the methods have low accuracy and are not quantitatively enough (or do not have a strictly linear relationship with seed aging time). More importantly, despite much research, there is still a very limited understanding of mechanisms underlying seed aging. Thus, there are great needs for understanding the molecular and biochemical basis of seed aging and for developing more precise and quantitative methods (or biomarkers) for assessing seed aging. In this project, we will investigate the changes of one type of macromolecules (seed stored mRNAs) and explore the potential of using them in determining seed aging status.

Before introducing the hypothesis and objectives, I will first review previous research on the main factors that have been suggested to play a role in seed aging and longevity.

1.2 Main factors affecting seed aging and longevity

1.2.1 Seed glassy state

It has been suggested and generally believed that orthodox seeds, which can survive desiccation and/or freezing during *ex situ* conservation, form a solid-like matrix with a high viscosity called the glassy state and such a state is believed to contribute to seed longevity greatly (Lima *et al*., 2017). The formation of the glassy state imposes stasis on intracellular reactions, limits molecular mobility, limits gel-liquid crystalline phase transformation of the lipid bilayer of cellular membranes, and prevents denaturation of macromolecules such as proteins (Bernal-Lugo and Leopold, 1998).

During maturation, seeds accumulate large amounts of sugars such as sucrose, raffinose, and stachyose, and the sugars are said to be one major factor for the formation of the glassy state (Amuti and Pollard, 1977; Bruni and Leopold, 1992; Koster, 1991). It is well known that sugars can interact with and protect macromolecules from denaturation and aggregation (Sun and Leopold, 1993). In addition, because of their high abundance, sugars may play a role in increasing the density of seed glasses through filling the free space between macromolecules, further restricting molecular mobility (Buitink and Leprince, 2004). However, the question regarding the extent that sugars contribute to the properties of

the seed glassy state is still unclear (Buitink and Leprince, 2004). The late embryogenesis abundant (LEA) proteins are also suggested as another major component for the formation of the glassy state (Wolkers *et al*., 1998). LEA proteins are known to accumulate highly during seed maturation, and suggested to protect macromolecules and fill in the space between macromolecules (Close, 1996). Since sugars and LEA proteins accumulate at the same time during maturation, their interactions may also contribute to the formation of the glassy state (Close, 1996; Gechev *et al*., 2012; Hoekstra *et al*., 2001). In addition, LEA proteins are also known for their roles in the tolerance of abiotic stresses such as chilling and salt stresses (Goyal *et al*., 2005). Other seed components may also have a role in forming glasses, including starch and cell wall materials such as pectin, hemicellulose, and cellulose (Buitink and Leprince, 2004).

In seeds, the glassy matrix can function in maintaining structural and enzymatic integrity. Protein structure in dry seeds appears to remain stable for a long period of time during natural aging (NA). It was reported that protein denaturation and aggregation were not observed in 28-year naturally aged seeds of onion, white cabbage, and radish (Golovina *et al*., 1997). Also, chemical and physical reactions are slow in orthodox seeds because of the limited molecular mobility. The density of the glassy state thus may affect seed aging (Wolkers *et al*., 1998). Since the intracellular viscosity and the glassy state could be enhanced by dry and cool conditions, the dry and cool conditions used for germplasm conservation in seedbanks would slow down damaging reactions and thus extend seed longevity (Buitink and Leprince, 2008). However, for recalcitrant seeds which do not form the glassy state and have high molecular mobility and active metabolism, continuous respiration and other cellular activities such as synthesis of macromolecules during storage could exhaust energetic reserves of those seeds faster (Berjak and Pammenter, 2008).

Although the glassy state contributes to seed cellular stability and longevity, the matrix itself may undergo time-dependent physical changes (Buitink and Leprince, 2008). Thus its physical properties and seed aging kinetics could also be affected leading to the accelerated seed viability loss after long-term storage (Walters, 2008).

1.2.2 Reactive oxygen species

Reactive oxygen species (ROS) refer to a set of highly reactive molecules that are derived from oxygen (Held, 2012). The oxygen atoms are prone to form radical species because of the two unpaired electrons in their outer electron shell. The addition of electrons to oxygen in the reduction reaction can give rise to a number of ROS such as superoxide (O_2^{-}·), hydrogen peroxide (H_2O_2), hydroxyl ion (HO^-), nitric oxide (·NO), and hydroxyl radical (·OH).

In living cells, ROS are produced mainly as byproducts during aerobic respiration through electron leakage of electron transport. In addition, many other biological processes may generate ROS. One instance is glycolate oxidases, which mainly function in the photorespiratory glycolate-recycling pathway and produce H_2O_2 as byproducts (Noctor *et al*., 2016). Other ROS-producing enzymes include amine oxidase, xanthine oxidase, and class III (guaiacol-type) peroxidase (Angelini *et al*., 2010; del Rio *et al*., 2006; Moschou *et al*., 2008; O'Brien *et al*., 2012). In Arabidopsis, there are over 70 genes have been identified to encode ROS-related genes (Cosio and Dunand, 2009). Such a large number of genes make it difficult to link the activity of one specific gene product to ROS production in cells. ROS are produced during the seed life cycle, from seed embryogenesis to maturation, desiccation, and storage as well as germination (Bailly *et al*., 2008). During seed storage, since most ROS producing enzymes are not active, non-enzymatic reactions such as autoxidation reactions， Maillard reactions (Sun and Leopold, 1995) and lipid peroxidation might be the main sources of ROS.

When ROS accumulate during seed storage, they may attack macromolecules continuously, resulting in damage and dysfunction of macromolecules, organelles, and various repair systems. In dry seeds, the glassy matrix would limit ROS mobility and their targets would be the adjacent molecules (Bailly *et al*., 2008). However, in imbibed seeds, free water would enable ROS (at least long-living H_2O_2), produced during dry storage and imbibition, to move within the seed and have a long-distance effect. If oxidative damage is severe, seeds would fail to initiate basic cellular metabolism and the germination process. It

has been reported that wheat seed aging is associated with an increased level of peroxide, and seed viability has a good correlation with the level of ROS-scavenging molecules (Spano *et al*., 2011).

Although ROS could cause various types of damage, ROS have important functions in seeds. The interaction between ROS production and scavenging could act as a messenger in seed dormancy release and germination (Bailly *et al*., 2008). This signaling function might be through the interplay between ROS and hormones, followed by the change in enzyme activity and/or gene expression. Also, ROS homoeostasis may be involved in regulating seed germination since it could perceive changes in environmental conditions (e.g. oxygen, temperature, and water availability) during seed germination (Bailly *et al*., 2008). Further, ROS may play a role in cell wall loosening, endosperm weakening, programmed cell death (PCD) of the aleuronic layer of seeds, and seedling protection against pathogens (Bailly *et al*., 2008). In consideration of both the damaging and beneficial effects of ROS accumulation in seeds, the concept of the oxidative window for seed viability was proposed (Bailly *et al*., 2008). According to this concept, the cellular processes associated with seed germination could occur within a permissive range of ROS levels.

In living cells, the ROS level is under tight control through ROS-scavenging systems which can be generally divided into two different types: enzymatic scavenging systems and non-enzymatic systems. The enzymatic system mainly consists of superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase (GSH-Px). SOD scavenges radical superoxide to H_2O_2 which is then removed by CAT. CAT, which mainly exists in peroxisomes, converts H_2O_2 into molecular oxygen and water. GSH-Px is mainly present in cytosol and mitochondria; it catalyzes the reduction of H_2O_2 to H_2O with glutathione (GSH) as the reducing molecule.

The components of the non-enzymatic system includes membrane-bound antioxidants such as tocopherol (vitamin E) and carotenoids (Kranner and Birtic, 2005), and water-soluble cytoplasmic antioxidants such as GSH (Bailly, 2004) and ascorbate (vitamin C) (Noctor and Foyer, 1998). Tocopherols and carotenoids accumulate highly on the plastid membranes, especially thylakoids; they are key compounds to scavenge singlet oxygen produced in light

harvesting (Noctor *et al*., 2016). The thiol group of GSH (a glutamyl-cysteinyl-glycine tripeptide) could function as a universal target of oxidative attacks (Kranner *et al*., 2006). As the most abundant low molecular-weight thiol molecule in plants, GSH could play numerous roles in maintaining cellular redox homeostasis, stress signaling, and ROS scavenging (Kranner *et al*., 2006). It also provides reducing power in regenerating a number of important antioxidants (e.g. tocopherol and ascorbate) (Foyer and Noctor, 2011; Pukacka and Ratajczak, 2007). The oxidized form of GSH (GSSG) can be reduced in a redox reaction catalyzed by the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent glutathione reductase. Since GSSG can be readily reduced to GSH comparing to its synthesis, the ratio of the oxidized form to the reduced form of GSH (GSSG/GSH) is widely accepted as an indicator for the redox state of an organism (Kranner *et al*., 2006). Ascorbate, like GSH, functions as a co-factor of several ROS scavenging enzymes in peroxide processing (Foyer and Noctor, 2011). For example, ascorbate acts as the reducing co-factor for ascorbate peroxidase, as well as in regenerating peroxidase (Noctor *et al*., 2016).

With little enzymatic activities under dehydration conditions, enzymatic antioxidants cannot react with ROS and antioxidants themselves could become inactive after prolonged storage. Furthermore, imbibition and germination of seeds can release ROS which are trapped in cellular components under the desiccation state and generate more ROS through metabolism. Several studies have shown that overexpressing enzymes in ROS scavenging could promote seed longevity and enhance stress tolerance. Seeds of tobacco plants overexpressing both Cu/Zn-superoxide dismutase (CuZnSOD) and ascorbate peroxidase (APX) genes show stronger antioxidant ability and more resistance to accelerated aging (AA) treatment than wild type seeds (Lee *et al*., 2010). Also, *Arabidopsis thaliana* plants overexpressing *NnMT2a* and *NnMT3,* which are members of metallothioneins involving ROS scavenging, also produce seeds more tolerant of aging treatment (Zhou *et al*., 2012).

1.2.3 Damage to lipids and membrane

Biological membranes are crucial for organisms in selective molecule transportation, cell signaling, biochemical reactions, and ATP production. One important component in the

biological membranes is polyunsaturated fatty acids (PUFAs). In living cells, PUFAs confer important properties to membranes such as selective permeability, flexibility, and fluidity.

Membrane lipids especially PUFAs are among the most vulnerable macromolecules to oxidation and damage (Reiss and Tappel, 1973). Both non-enzymatic and enzymatic reactions contribute to the damage on membrane lipids (Shewfelt and Purvis, 1995). In non-enzymatic PUFA decomposition, a hydrogen moiety of the unsaturated carbon from PUFAs could be captured by a free radical to form H_2O_2 (Held, 2012). The consequent unpaired electron on the PUFA can then capture an electron from an oxygen, resulting in a peroxy radical. Lipid peroxides are short-lived and could decompose to produce a series of molecules, including reactive carbonyl compounds (RCCs) (Held, 2012), 4-hydroxyhexenal, and 4-hydroxynonenal, all of which are highly reactive (Negre-Salvayre *et al*., 2008). This chain reaction is called lipid peroxidation. The by-products of lipid peroxidation have a longer lifespan than ROS and can react with proteins and DNA resulting in their oxidation. As a result, membrane properties including selective permeability, fluidity, and activities of membrane-bound enzymes can be affected. In addition, the changes in membrane permeability and fluidity can affect seeds by disrupting the glassy state. The oxidation and damage to membrane PUFAs and phospholipids by peroxidation are widely regarded as one of the major causes for seed aging during storage (Wilson and Mcdonald, 1986).

The membrane can be modified through enzymatic reactions, for instance through lipoxygenases. Lipoxygenases belong to the family of nonheme iron-containing dioxygenase, and they catalyze the oxygenation of PUFAs and produce conjugated diene hydroperoxides (Loiseau *et al*., 2001). Although lipoxygenases are present in seeds of many plants, their involvement in lipid peroxidation during seed aging is still unknown. It has been shown that the products of PUFA peroxidation are lower in the *lipoxygenase-3* (*lox-3*) mutant which is deficient in LOX3 activity (Suzuki *et al*., 1999; Suzuki *et al*., 1996), but it is unknown whether the mutant seeds age differently from the wild type seeds. Another example is lipoxygenase 2 (LOX2) which also catalyzes the oxidation of PUFAs during seed germination to produce unsaturated fatty acid hydroperoxides. Overexpression of *LOX2* in rice (*oryza sativa* L.) resulted in a reduced germination percentage when seeds were

subjected to accelerated aging, while down-regulation of *LOX2* by RNAi increased seed longevity (Bueso *et al*., 2014). Phospholipase D (PLD) is also assumed to be responsible for membrane disruption and the seed aging process (Devaiah *et al*., 2007). PLDα1, the major form of PLD, has been reported as a player in ROS production, lipid peroxidation, and seed aging. The Arabidopsis knockout mutant deficient in PLDa1 activity produced fewer lipid peroxides and maintained a higher level of PUFAs comparing to wild type seeds during seed aging (Devaiah *et al*., 2007). The knockout mutant seeds showed more tolerance to seed aging treatments, providing support for a role of PLD in membrane disruption and seed aging.

For cellular membranes, a major free radical scavenger is vitamin E. Because of its lipophilic property, vitamin E can exist in the lipid bilayer of membranes (Pamplona, 2008). Vitamin E can reduce lipid peroxyl groups to hydroperoxides while being oxidized and thus stop the lipid peroxidation process (Pamplona, 2008). It has been shown that Arabidopsis mutants that are defective in vitamin E synthesis produce seeds with reduced longevity (Sattler *et al*., 2004).

1.2.4 Mitochondrial damage

While mitochondria in cells produce energy, they also produce ROS and can be damaged by the self-produced ROS. MtDNA (mitochondrial DNA) is more sensitive to oxidative stress than gDNA (genomic DNA) for the following reasons. First, mtDNA, unlike gDNA, is not protected by histones. Although protein-DNA complexes (nucleoids) in mitochondria play a similar role as histones, they are less effective (Sastre *et al*., 2000). Second, mtDNA is closer to the source of ROS production. Third, there are no introns in mtDNA and any damage on mtDNA may affect protein expression (Johns, 1995). There is some evidence that mtDNA damage accumulate during seed aging (Gadaleta *et al*., 1990). It has been reported that the decline in the capacity of oxidative phosphorylation is correlated with the accumulation of mtDNA base deletions during seed aging (Lezza *et al*., 1994). Thus, mutations in mtDNA have been suggested to be a contributing factor to seed aging.

Mitochondrial GSH is a major factor regulating oxidative status in mitochondria.

During the aging process, severe oxidation of mitochondrial GSH induces an imbalance in the redox state which is correlated to seed aging (de la Asuncion *et al*., 1996). Thus, maintaining an adequate level of the reduced GSH may be important for reducing mtDNA damage and seed survival.

Mitochondrion-specific acylhydrolase can protect the mitochondrial membrane from damage. Transgenic seeds overexpressing *AtDLAH* (*Arabidopsis thaliana* DAD1–like acylhydrolase) showed an enhanced tolerance to seed aging treatments while seeds of a knockdown mutant showed reduced germination performance after aging treatments. In addition, the seeds with enhanced longevity showed a lower level of lipid peroxidation and higher level of polar lipids than the wild type seeds (Seo *et al*., 2011), suggesting a link between damage to mitochondrial membrane and seed aging.

1.2.5 Damage to proteins

During maturation, seeds accumulate several classes of proteins, including globulins, prolamins, albumins, and glutelins, which may play fundamental roles in seed resistance to desiccation, aging, and imbibition damage (Ibl and Stoger, 2012; Bojórquez-Velázquez *et al*., 2019; Farrant *et al*., 1992). Some stored enzymes also function to restart basic cellular metabolism in the early stage of seed germination. These enzymes may not be functional in dry seeds, but with increased seed water content during rehydration, enzyme activities can be restored quickly (McDonald, 1999). Even in the absence of *de novo* protein synthesis, some important biological processes like translation and transcription can restart quickly with stored enzymes, and thus stored proteins may be critical at the early stage of germination (Rajjou and Debeaujon, 2008; Rajjou *et al*., 2004).

Because of their high abundance, stored proteins can be the primary targets of ROS. Sulfur-containing amino acids are among the more vulnerable residues to be oxidized, which could result in enzyme deactivation. In some cases, oxidation of sulfur-containing residues and the enzyme deactivation can be reversed. For example, disulfide bridges could form between two cysteine residues and be reversed during seed germination accompanied by the oxidation of thioredoxins (Job *et al*., 2005). In contrast to reversible protein modifications,

irreversible oxidative modifications also occur during seed storage. Several mechanisms can result in protein oxidation, such as carbonylation of amino acid residues, nitration of tyrosine, and glycoxidation (Berlett and Stadtman, 1997; Davies, 2005; Dukan *et al*., 2000; Dunlop *et al*., 2002; Rivett and Levine, 1990), with carbonylation being the major one (Møller *et al*., 2011).

Protein oxidation is not necessarily detrimental to plant seeds, since it has been reported that protein oxidation during seed aging could play several roles in seed germination [as reviewed by (El-Maarouf-Bouteau *et al*., 2013)]. These roles could include (1) ROS scavenging. Since proteins are abundantly present in seeds, they could buffer the damaging effects from continuously produced ROS during seed storage and germination, (2) seed dormancy releasing. Carbonylation of certain proteins might promote the release of seed dormancy (Oracz *et al*., 2007), (3) protein mobilization: carbonylated proteins are more vulnerable to proteolysis and can be easily hydrolyzed into amino acids during seed germination, and (4) enzyme regulation: oxidation of the proteome may result in a wide range of enzyme deactivation and contribute to metabolic quiescence in mature dry seeds, but they may be re-activated through the reduction of disulfide bridges (Alkhalfioui *et al*., 2007). Since seeds need to switch basic metabolic processes off in dry storage and on during rehydration, enzyme deactivation and reactivation through oxidative modification could be a simple and efficient way (Holdsworth *et al*., 2008).

Since the damage occurs to proteins in seed storage, repairing the damage will help seed survival. However, accumulated protein damage during seed storage cannot be repaired in dry seed state (Chatelain *et al*., 2013). In the early stage of seed germination, when seed water content increases and enzyme activities are restored, the protein damage accumulated during seed storage can thus be repaired (McDonald, 1999). One of the repair pathways on damaged proteins, the repair of aspartate/asparagine by protein-isoaspartyl-methytransferase (PIMT), appears to be important for seed longevity and successful germination (Mudgett *et al*., 1997; Oge *et al*., 2008). The deamination of Asn or dehydration of Asp could result in a succinimidyl ring, which is unstable, and its breakdown can result in a mixture of Asp and abnormal isoAsp residues (Aswad *et al*., 2000; Geiger and Clarke, 1987). The isoAsp could

affect protein folding and enzyme activity (Esposito *et al*., 2000; Mamula *et al*., 1999). The isoAsp residues can be recognized and repaired by PIMT specifically (Aswad *et al*., 2000). In PIMT-deficient Arabidopsis mutants, seeds showed a higher level of isoAsp content and a faster speed of seed aging compard to the wild type (Oge *et al*., 2008). These results provide some evidence to show that the repair of oxidized proteins could benefit seeds in long-term storage and seedling establishment (Oge *et al*., 2008).

1.2.6 Damage to DNA

Seed DNA is susceptible to endogenous and exogenous damage. For example, nucleic acid bases can be damaged by free radicals and alkylating agents (Doetsch, 1995; Plooy *et al*., 1984). The damage to DNA could be covalent modifications on sugar moieties and nucleobases, single strand breaks, and changes in DNA structure or content (Cadet *et al*., 2012). Moreover, double-strand breaks could happen if the random nicks on the two DNA strands have a distance of fewer than 15 bps (Ventura *et al*., 2012). DNA laddering on an agarose gel, indicative of the occurrence of double-strand breaks, is one of the most observed patterns of DNA damage in aged seeds. It has been shown that loss of seed viability during storage is accompanied by the gradual degradation of DNA into fragments (Kranner *et al*., 2011).

For dry seeds under storage, free radicals and lipid peroxidation have been suggested to be mainly responsible for nucleic acid lesions. As discussed earlier, free radicals accumulate during seed storage and can react with DNA. For example, the reactive hydroxyl radical could attack DNA on the 2-deoxyribose and nucleobase moieties. Other ROS-related attacks on cytosine, thymine, adenine, thymine, 5-mythylcytosine, and guanine could result in base modifications. Among the derivatives, the most common product of nucleic acid oxidation by free radicals is reported to be 7, 8-dihydro-8-oxoguanine (8-oxo-G) (Cadet *et al*., 2012; Ventura *et al*., 2012; Yang *et al*., 2001). The 8-oxo-G is mutagenic since it can mispair with an adenine instead of a cytosine during DNA replication. DNA damage can also be caused by lipid peroxidation. Lipid peroxidation produces a variety of nucleophiles that can react with nucleic acids and proteins (Marnett, 1999). For example, malondialdehyde,

one of lipid peroxidation products, reacts with G residues in nucleic acids, leading to the formation of pyrimidopurinone (M_1G) which can block DNA base pairing (Marnett, 1999). Thus, DNA damage could affect DNA replication during the germination process (Painter, 1985; Protic-Sabljic and Kraemer, 1985).

To maintain the integrity of the genome, both prokaryotic and eukaryotic organisms possess various mechanisms to repair DNA damage, including direct reversal, base excision repair, nucleotide excision repair, double-strand break repair, and mismatch repair pathways (Britt, 1999). There are specific reviews on DNA repair systems in plants (Chatelain *et al*., 2012; Tuteja *et al*., 2001; Ventura *et al*., 2012). Detailed discussion on DNA repair is beyond this background review on seed longevity. Briefly, while the direct reversal pathway involves single proteins, the other repair pathways require multiple proteins to repair DNA damage. The nucleotide excision repair excises a patch of 25-32 nucleotide oligomer including the lesion while the base excision repair pathway repairs single defective base. Either homologous recombination or non-homologous end joining could be used for double-strand break repairs. One example of DNA repair in seeds was reported by Chen *et al*., (2012). As indicated earlier, one common product of DNA damage caused by ROS is 8-oxo-G which leads to transversion from GC to TA in the subsequent DNA replication (Yoshida *et al*., 2002). The 8-oxo-G on DNA could be repaired by 8-oxoguanine-DNA glycosylase 1 (OGG1) which functions as glycosylase/ apurinic/ apyrimidinic (AP) lyase to excise the damaged base (Nash *et al*., 1996). It is interesting to note that overexpressing *OGG1* in *Arabidopsis thaliana* decreased the level of 8-oxo-G and rendered the seeds more resistant to seed aging treatment, indicating that DNA damage contribute to seed deterioration (Chen *et al*., 2012).

1.2.7 Stored mRNAs and seed aging

1.2.7.1. Seed stored mRNAs

Mature plant seeds contain various mRNAs referred to as stored mRNAs (Dure and Waters, 1965; Marcus and Feeley, 1964), and also as long-lived mRNAs because they are present in seeds from late embryogenesis to early seed germination (Sano *et al*., 2015). It is

assumed that stored mRNA accumulation is controlled by the interplay between endogenous (e.g. hormones) and exogenous factors in seed development and maturation (Nakabayashi *et al*., 2005). In one study using Arabidopsis seeds, over 12,000 mRNA species were detected and different varieties had very similar long-lived mRNA profiles (Kimura and Nambara, 2010). In addition, the classes of long-lived mRNAs are highly conserved between dicot Arabidopsis and monocot barley seeds (Rajjou *et al*., 2012), suggesting the functional importance of seed stored mRNAs.

One of the major classes of abundant stored mRNAs is from the LEA genes (Kimura and Nambara, 2010). Among Arabidopsis stored mRNA, 17 out of 51 LEA genes were found at a high level in dry seeds (Kimura and Nambara, 2010). Another major class of stored mRNAs is the templates for seed stored proteins, especially for three cruciferins (12 S globulins) and two napins (2S albumins) (Fujiwara, 2002). The third abundant group of long-lived mRNAs is from genes involved in stress response and tolerance such as heat shock proteins (Kimura and Nambara, 2010).

1.2.7.2 Roles of seed stored mRNAs

The most fundamental and plausible role for stored mRNA may be to provide templates for protein synthesis during seed germination. The stored mRNAs encode many proteins of diverse processes (Kimura and Nambara, 2010; Nakabayashi *et al*., 2005; Sano *et al*., 2015; Sano *et al*., 2012) and may be needed for protein synthesis during the early stages of seed germination (Comai *et al*., 1989; Rajjou *et al*., 2004; Sano *et al*., 2015). When germinating Arabidopsis seeds were treated with α-amanitin, a potent inhibitor of DNA-dependent RNA polymerase II, seeds could still germinate successfully although the seed vigor was severely affected. However, they they completely failed to germinate when treated with cycloheximide, a protein translation inhibitor, suggesting that stored mRNAs could support initial seed germination without the synthesis of new mRNAs (Rajjou *et al*., 2004). Sano *et al.* showed that in rice stored mRNAs increased from 10 days after flowering (DAF) and became highly abundant at 40 DAF (Sano *et al*., 2015). While the germination of 10 DAF embryos was severely impaired by the transcriptional inhibitor actinomycin D (ActD),

over 70% of embryos of 20 – 40 DAF could germinate in the presence of ActD, supporting the idea that accumulation of stored mRNAs is critical for seed germination and the resistance to ActD inhibition (Sano *et al*., 2015).

Based on the observation on stored mRNAs, Rajjou *et al*. concluded that the seed germination progress has already been programmed and stored in seed materials during seed maturation (Rajjou and Debeaujon, 2008). It is reasonable to think that the stored mRNAs may encode critical enzymes required for restarting cellular metabolism after seed rehydration and without them seed germination could be impaired. Although stored mRNAs seem to be a prerequisite for successful germination, only a portion of them are selectively translated into proteins during seed germination while the others may be residue templates for synthesizing storage proteins during seed development (Galland *et al*., 2014; Galland and Rajjou, 2015; Rajjou and Debeaujon, 2008).

In addition to the selective translation, selective degradation of stored mRNA is also an important process in regulating gene-selective expression (Leymarie *et al*., 2012). Since degrading an mRNA is faster than synthesizing it, it has been proposed that mRNA degradation can serve as an effective mechanism in fine-tuning gene expression in seed germination (Basbouss-Serhal *et al*., 2017). mRNAs could be degraded through either 5' or 3' end. The exosome organelle has multi-subunit complexes for 3' to 5' RNA degeneration (Januszyk and Lima, 2014). The nucleus exosome could process different types of RNA, including rRNA, snRNA, mRNA, and tRNA (Lange *et al*., 2014). The cytoplasm exosome plays roles in mRNA degradation and mRNA surveillance such as nonsense-mediated decay (NMD) (Chlebowski *et al*., 2013). mRNA degradation can be initialized by polyA shortening, catalyzed by multi-subunit carbon catabolite repressor 4-negative on TATA complex (CCR4-NOT), so-called the deadenylation process (Chen *et al*., 2002). In 5'-3' mRNA degradation, the eukaryotic conserved decapping complex C can catalyze mRNA 5' decapping after 3' deadenylation, followed by degradation in the cytoplasm (Xu *et al*., 2006). In this regard, it is interesting to note that 5' to 3' selective mRNA decay may have a role in seed germination (Leymarie *et al*., 2012).

1.2.7.3 Damage to seed stored mRNAs

In dry seeds, RNA bases can be modified or damaged by oxidative reagents during storage. Under the oxidative environment, the most abundant base modification is 8-oxo-G (Barciszewski *et al*., 1999; El-Maarouf-Bouteau *et al*., 2013). In the production of 8-oxo-G, a 3-hydroxyoctanoylcarnitine (C8-OH) adduct radical formed firstly from the OH· attack on guanine. Subsequently, the unstable C8-OH loses one electron and proton, and finally produces 8-oxo-G. Since 8-oxo-G is relatively easy to be identified and quantified with HPLC technique, it has been used as a marker for mRNA oxidation (El-Maarouf-Bouteau *et al*., 2013). In a pioneering study on sunflower seeds, mRNA oxidation was observed during seed dormancy release, reflected by a 50% increase in the amount of 8-oxo-G (Bazin *et al*., 2011).

In the germination process, the modified or damaged mRNA bases could negatively affect protein translation, resulting in mRNA stalling on the ribosome and/or reduced/truncated protein production (Ventura *et al*., 2012). In living cells, cytoplasm processing bodies (P-bodies), which refers to the distinct foci containing the enzymes involved in mRNA turnover (Bashkirov *et al*., 1997), could play roles in degrading the oxidized mRNA with its 5' to 3' exonuclease, decapping enzymes, and activators of decapping (Sheth and Parker, 2006). Interestingly, ROS could induce P-body formation in the cytoplasm (Shan *et al*., 2007). In addition, in severely aged wheat seeds, its ability to degrade stored mRNA and to *de novo* synthesize mRNA were reduced significantly (Rushton and Bray, 1987). It is also suggested that there is a correlation between seed germination percentage and the ability to synthesize mRNA *de novo* in wheat seeds (Rushton and Bray, 1987).

In general, successful seed germination may be controlled by the balance among mRNA accumulation, degradation, sequestration, protection, and recruitment in protein translation (Galland *et al*., 2014). Although the specific mechanisms remain to be fully understood, experimental evidence has so far supported the generally accepted concept that stored mRNAs are important for early stages of seed germination. It is thus reasonable to

suggest that stored mRNAs and particularly their degradation are important for seed aging. Prior to the start of this project, little research had been done to determine the degradation of stored mRNAs during plant seed aging process. More specifically discussion on stored mRNA degradation and seed aging can be found in Chapters 3 and 4.

1.2.8 Pathways of mRNA degradation

Although my study is on mRNA degradation in seeds, it is useful to review mRNA degradation in normal cells. mRNA degradation is important for controlling mRNA level and quality in eukaryotes including plants. Several major pathways for mRNA degradation are known: 1) 3'-5' mRNA degradation pathway (Chlebowski et al., 2013), 2) 5'-3' mRNA degradation pathway (Nagarajan et al., 2013), 3) endonuclease cleavage pathway (Hammond, 2005), and 4) mRNA surveillance pathway including nonsense-mediated decay (NMD), non-stop decay (NSD) and no-go decay (NGD) respectively (Shoemaker and Green, 2012;Łabno et al., 2016). These pathways may cooperate with each other to complete mRNA degradation.

Nearly all mRNAs contain a stretch of adenine bases at the 3' end, which is referred to as the poly(A) tail. Removal of the poly(A) tail (or deadenylation) is the initial step in the 3'-5' or 5'-3' mRNA degradation pathway. The deadenylation is achieved mainly through the poly(A) ribonuclease (PAN) 2/PAN3 and carbon catabolite repressor 4- Negative on TATA (CCR4-NOT) multi-protein complexes in eukaryotes (Wolf and Passmore, 2014). In Arabidopsis, PAN and CCR4-NOT complex have been analyzed genetically. It has been reported that AtPAN was an important player in seed germination and the mutant seeds were very sensitive to exogenous ABA and salicylic acid (Nishimura et al., 2005). In addition, Arabidopsis CCR4-associate-factor 1A (AtCAF1A) and AtCAF1B were reported to participate in cellular defense response to biotic and abiotic stresses (Liang et al., 2009;Walley et al., 2010).

Poly(A) shortening might be followed by the 5' decapping process (Beelman and Parker, 1995). The cap structure protects the mRNA against 5'-3' exoribonucleases, and decapping is required prior to 5'-3' mRNA degradation. The decapping process is achieved

through decapping complex which consists of Decapping Protein 1 (DCP1) (LaGrandeur and Parker, 1998;Dunckley and Parker, 1999) and DCP2 (Lykke-Andersen, 2002;van Dijk et al., 2011). After decapping, the uncapped mRNA is degraded by the 5'-3' exoribonuclease mainly in cytoplasmic foci called processing bodies (Goeres et al., 2007;Weber et al., 2008;Hamada et al., 2012;Nagarajan et al., 2013;Motomura et al., 2014). The 5'-3' mRNA degradation plays an important role in stress response in plants. For example, under normal growth conditions (22ºC), AtDCP1 accumulates and forms AtDCP1-bodies while AtDCP2 disperses in the cytoplasm. However, AtDCP2 starts to accumulate together with AtDCP1 and forms AtDCP1/AtDCP2-bodies (Motomura et al., 2014). Moreover, the defect in the 5'-3' mRNA degradation pathway could result in seedling death as observed in Arabidopsis atdcp1 or atdcp2 null mutants (Xu et al., 2006;Goeres et al., 2007;Iwasaki et al., 2007;Motomura et al., 2012).

In addition to mRNA degradation catalyzed by 3'-5' and 5'-3' exoribonucleases, mRNA can also be mediated by endonucleases. The Argonaute (AGO) is one prominent class of endonucleases present in diverse species and AtAGO1 plays a major role in plant mRNA turnover. AGO could bind different classes of small non-coding RNAs, such as microRNA (miRNA) and small interfering RNA (siRNA) and form RNA-induced silencing complex (RISC) (Yoo et al., 2011). The AGO of the complex is guided by the small RNAs to and cut the target mRNAs based on sequence complementarity (Yoo et al., 2011). The target mRNAs are then completed degraded through the 5'-3' and 3'-5' pathways (Houseley and Tollervey, 2009). It was reported that miRNA/siRNA play various roles in plant development. For instance, the miR159/319 could affect leaf morphology through targeting the transcription factor genes TCP and MYB (Palatnik et al., 2003;Allen et al., 2007). TAS3-siRNAs triggers mRNA cleavage of the auxin response factors ARF3 and ARF4, which are important for the specification of leaf polarity (Garcia et al., 2006).

In eukaryotes, there are surveillance pathways to recognize aberrant mRNAs, such as those with a premature stop codon (PTC) or lacking stop codon, so that they are unlikely to be translated into incorrect proteins. The known pathways in mRNA surveillance include: nonsense-mediated decay (NMD), non-stop decay (NSD) and no-go decay (NGD) (Łabno et

al., 2016) pathways. The most well-studied mRNA surveillance pathway is the NMD pathway. After mRNA splicing, the exon-exon junction complex (EJC) binds to the conjunction sites on mature mRNA and will be removed after the first round of translation by the ribosomes (Zhang and Guo, 2017). If any of these EJC complexes remains bound to the mRNA template after the first round of translation, NMD is activated (Dai et al., 2016). mRNA decay is then initiated by the removal of 5' cap followed by 5'-3' degradation or by the deadenylation from 3'-5' (Dai et al., 2016). The NMD pathway has been reported to play an important role in plant adaptation to environmental changes and plant defense response (Chen and Shyu, 2011). In Arabidopsis, many PTC-containing transcripts resulting from alternative splicing were found, and the aberrant transcripts were degraded by NMD pathway (Drechsel et al., 2013). The environmental changes such as extreme temperatures, salt stress, and photoperiod changes could result in an increased level in the aberrant mRNA transcripts of the Flowering Locus M (FLM), Early Flowering 3 (ELF3) and Timing of Cab Expression 1 and these aberrant mRNAs were degraded through NMD (Kwon et al., 2014;Sureshkumar et al., 2016). Apart from NMD, little is known about the other two mRNA surveillance pathways in plants. In general, NSD is considered to be for degrading mRNAs with no stop codon (Frischmeyer et al., 2002;van Hoof et al., 2002) while NGD deals with mRNAs with the sequence that could stall ribosomes on the mRNA strand during protein translation.

1.2.9 Protective compounds or biological processes

A range of protective compounds may also contribute to seed longevity (Leprince *et al*., 2017; Sano *et al*., 2016). Orthodox seeds accumulate a large amount of LEA proteins, heat shock proteins (HSPs) (Tejedor-Cano *et al*., 2010) as well as non-reducing soluble sugars (the sucrose and raffinose family of oligosaccharides) (Salvi *et al*., 2016). It has been proposed that those molecules could reduce the rate of water loss in seed drying and provide an opportunity for cells to cope with the mechanical shrinking. Also, their polar groups may interact with the surface of macromolecules and replace water in stabilizing structures of the macromolecules, which is referred to as the "water replacement hypothesis" (Crowe and Crowe, 1992).

Seed longevity could also be enhanced by antioxidant molecules, such as tocopherol (Sattler *et al*., 2004), glutathione (Nagel *et al*., 2015) and flavonoids (Debeaujon *et al*., 2000). In addition, DNA and protein repair systems (Oge *et al*., 2008; Waterworth *et al*., 2010) promote the metabolic restart during seed germination through fixing damage on the macromolecules, and thus could improve seed viability.

Other mechanisms in extending seed longevity include: 1) hard seed coat which could prevent air and water penetration, minimize ROS production inside seeds and reduce mechanical damage during imbibition (Debeaujon *et al*., 2000); 2) interactions of temperature, relative humidity and unknown factors, which could impact the rate of deteriorative reactions (Walters *et al*., 2005); and 3) the availability of stored macromolecules, such DNA, RNA, and protein, which are needed to restart the basic metabolism following imbibition (Rajjou and Debeaujon, 2008; Sano *et al*., 2016).

1.2.10 Seed coat

The seed coat refers to the tissue surrounding the embryo and nutritive materials; and it is developed from the chalazal and integuments of the ovule. The seed coat could play many roles, such as ROS scavenging and mechanical protection. The flavonoids in the seed coat could function as antioxidants and scavenge ROS in seed aging (Weisshaar and Jenkins, 1998). Compared to wild type seeds, mutant seeds with a defective seed coat, have a lower germination percentage after accelerated seed aging or long-term storage (Appelhagen *et al*., 2011). Flavonoids such as proanthocyanidins might also function as a physical barrier against fungi infection, since soluble proanthocyanidins could be oxidized into insoluble polymers and quinonic products, both of which cling to cell walls and form a chemical-physical barrier (Mayer and Staples, 2002; Pourcel *et al*., 2005). The resultant quinone barrier could enhance the mechanical strength of the seed coat and thus reduce mechanical damage (Pourcel *et al*., 2005). Seed coats of some species also contain lignin and its accumulation may contribute to seed resistance to mechanical damage (Sano *et al*., 2016). During imbibition, the seed coat restricts the rate of water absorption and solute leakage to the water, alleviating imbibitional damage (Crowe *et al*., 1987).

Although physically relatively strong compared to other components of a seeds, the seed coat could deteriorate gradually during the seed aging process, and a damaged seed coat would render seed more sensitive to sub-optimal germination conditions.

1.2.11 Typical methods in monitoring seed aging

Currently, there are no perfect seed monitoring tools that can save seeds and provide precise aging information at the same time. The seed aging status is commonly determined through seed germination tests. Usually, germination tests are performed under laboratory conditions which are optimal for seedling establishment. However, field conditions may differ largely from the laboratory environments, and thus bench tests on seed germination and vigor may not predict its field performance properly (Corbineau, 2012). Also Fu *et al*. (2017) showed that the traditional germination tests were not so accurate in assaying more than 500 accessions of wheat seeds. As alternatives, shoot elongation (rye-grass, maize, wheat, barley, and oat) and seedling weight (maize and soybean) are used to predict seed field performance (Corbineau, 2012). Another method in monitoring seed viability is the topographical tetrazolium test (Yaklich and Kulik, 1979). This method takes advantage of dehydrogenase activity, which exists in living tissues and catalyzes the reduction of the colorless tetrazolium salt into a red non-diffusible dye. Thus all living tissue in the seed embryo can be stained to red by tetrazolium salt, enabling to determine the seeds viability or otherwise. Other methods or markers have also been developed based on various aspects of seed physiology. These include ethylene production, electrolyte leakage during seed imbibition, the ROS scavenging ability of antioxidant system (e.g. CAT, SOD, GSH), sugar metabolism (e.g. changes in the raffinose family) and other markers (e.g. beta-tubulin and DNA replication) (Fu *et al*., 2015). These markers have potential in evaluating seed maturity, seed quality, seed aging, and longevity. However, some of these methods are not easy to perform while some other markers are not very accurate in evaluating seed viability. Further, a precise evaluation of seed quality often requires more than one marker, and different markers may be differently suited for different species (Corbineau, 2012). A better understanding of seed longevity, deterioration, and germination at the biochemical, molecular and cellular levels are important,

not only for understanding the fundamental aspects of seed aging, but also for developing more precise and effective methods of evaluating seed aging status.

1.3 Research questions, hypothesis, and objectives

To better understand seed aging phenomenon, to find better biomarkers of seed deterioration, and to facilitate germplasm conservation and seedling establishment in agricultural production, it is important to identify the critical factors of seed aging and understand how they work mechanistically. It can be envisioned that the contributions by different factors may vary in different plant species, as often encountered with complex biological phenomena. Thus, in this project initial efforts will be focused on experimentally tractable model systems such as Arabidopsis where considerable knowledge has accumulated. The findings on Arabidopsis could be further applied to crop species such as canola and wheat.

As discussed earlier, seed stored mRNAs seem to play a pivotal role and is one of the most vulnerable factors in seed aging. Prior to the start of this project, there has been little research examining the changes of stored mRNAs during seed aging, except for those reporting the degradation of ribosomal RNAs. Earlier on, it was observed that the amount and integrity of rRNAs decreased in non-viable embryos of rye grains (Roberts *et al*., 1973) and pea embryonic axis tissues (Bray and Chow, 1976). Deterioration in rRNA integrity was observed in seeds of plants such as carrot, tobacco, sunflower, and soybean, based on the appearance of shorter RNA fragments supposedly due to the degradation of 25S and 18S rRNAs (Brocklehurst and Fraser, 1980; Reuzeau and Cavalie, 1997; Thompson *et al*., 1987) or on the relative amount of intact 25S and 18S rRNAs (Brocklehurst and Fraser, 1980; Fleming *et al*., 2017; Kranner *et al*., 2006).

Since mRNAs, in general, are prone to degradation, we postulated that seed stored mRNAs would undergo degradation during seed aging. Specifically, we hypothesized that seed stored mRNAs are gradually damaged during seed storage. If so, the damage or degradation should be detectable and further, stored mRNAs may be used as markers of seed aging.

The specific objectives of this study are:

1) With Arabidopsis as a model, determine stored mRNA levels and their degradation in aged seeds

2) Develop quantitative methods to analyze and quantify the changes in mRNA degradation during seed aging and determine the relationship of stored mRNA degradation and seed aging time,

3) Based on stored mRNA degradation, develop a method for monitoring seed aging, and compare the method with traditional methods,

4) Compare stored mRNA degradation in naturally aged (NA) and accelerated-aged (AA) Arabidopsis seeds, and

5) Perform similar research in important crops such as canola and wheat and compare the findings from different plants.

2. MATERIALS AND METHODS

2.1 Seed materials

For *Arabidopsis thaliana*, the ecotype "Columbia" was used in this study. To have uniform germination, seed priming was conducted before sowing seeds. About 300 seeds were placed in a 2 mL microtube filled with 1 mL water. The tube was stored in a 4-5 °C refrigerator for 2 days. Seeds were sown in soil (Sunshine Mix #3, Univ. of Saskatchewan Phytotron) in pots (4 x 4 inches) which were placed in a growth room (20 °C constant, 16/8 h day/night cycles with a daylight fluorescent white light of 100 ± 15 μmoles m^{-2} sec^{-1}). Arabidopsis inflorescence is indeterminate and new flowers are being produced continuously. When the siliques from the early flowers started to desiccate, the new flowers being produced on the top of inflorescence branches were removed so that the siliques could mature fully and desiccate to obtain seeds with consistent quality. After harvest, the seeds were left drying in normal room conditions for 3 days. Dried seeds were stored in 2 mL microtubes with airproof screw caps in a 4~5 ^{o}C refrigerator.

For wheat, seeds of *Triticum aestivum* cultivar 'Superb' (kindly provided by Carolee Horbach and Gregory Peterson, Saskatoon Research and Development Centre, AAFC) were harvested in 2015, and were sealed and stored in an airtight aluminum bag in a 4~5 ^{o}C refrigerator.

For canola, seeds of *Brassica napus* cultivar 'Westar' (kindly provided by Hao Hu, Saskatoon Research and Development Centre, AAFC) were harvested in 2017, and were sealed and stored in an airtight aluminum bag in a refrigerator.

2.2 Seed aging treatments

For Arabidopsis, seeds have been collected in Dr. Wang's lab over the years and stored in a 4 °C refrigerator (Table 2.1). These seeds were used as naturally aged (NA) seeds. The accelerated aging (AA) treatments of Arabidopsis seeds were performed using the conditions similar to those described previously (Sugliani *et al*., 2009). Dry seeds were placed in 2 mL tubes with the cap removed (each tube having 100 mg seeds). The tubes were

placed into a plastic container with a saturated KCl solution at the bottom. The container was closed with a lid, sealed with paraffin, and placed in an incubator for aging treatments at 37 °C and other temperatures. Following the treatment, the seed samples were removed, air-dried for 3 days at room temperature and then stored at 4-5 °C with the tube cap closed. Usually, the same seed source was used for AA treatments in a series of experiments for comparing different treatments.

For AA treatments on wheat and canola seeds, the seeds were much larger and a 2 mL tube could not hold many seeds. Therefore we set up a treatment chamber using a plastic container box. Briefly, one thin layer of dry wheat seeds was distributed on the bottom of a mesh holder. The mesh holder was placed into a plastic container with four bottles of saturated KCl solution at the bottom to create high humidity of about 82% at 37 to 40 °C. The plastic container was capped with a lid, sealed with plastic wrap, and placed into an incubator (40 °C for wheat and 37 °C for canola). Seed samples (180 wheat seeds, or 350 canola seeds for each AA treatment) were incubated for specific lengths of time, removed from the container, air-dried for 3 days at room temperature and then stored at 4-5 °C until use. The same seed source for wheat and canola was used for AA treatments.

Table 2.1 List of naturally aged Arabidopsis seeds used in this study.

Seed lot	Harvest time	Seed storage time (years) [(1)]
WT	Sep.2014	0
854	Mar.2010	5
357	Sep.2003	11.5
290	Dec.1999	15
51	Dec.1997	17
26-14	Dec.1997	17
28-16	Dec.1997	17
29-17	Dec.1997	17

(1) The seed storage time refers to the time of harvest to the time when the germination assay was performed in Mar. 2015.
The harvested seeds were were stored in 2 mL microtubes with airproof screw caps in a 4 ^{o}C refrigerator.

2.3 Isolation of total RNA from seeds

Total RNA was isolated from dry seeds as described (Onate-Sanchez and Vicente-Carbajosa, 2008) with modifications. In brief, 45 mg Arabidopsis seeds were frozen in liquid nitrogen and ground with mortar and pestle into powder. The powder was added to a mixture of 550 μL extraction buffer [0.4 M LiCl, 0.2 M Tris (pH 8.0), 25 mM EDTA, 1% SDS] and 550 μL chloroform in a 1.5 mL tube. The content was well mixed with handshaking and vortexing (~10 sec vortexing). Following 3-min centrifugation, the upper phase (500 μl) was transferred to a new 1.5 mL tube containing 500 μL phenol. After brief vortexing (~10 sec) and 5 min incubation at room temperature, 200 μL chloroform was added, which was followed by 3 min centrifugation. The upper phase (500 μl) was transferred into a new 1.5 mL tube containing 170 μL 8 M LiCl. After 30 min storage in a -20 °C freezer, the mixture was centrifuged for 15 min at 4 °C. The supernatant was discarded and the pellet was further treated with DNase I (Roche). The DNase I treated mixture was transferred into a solution containing 500 μL DEPC-H_2O, 250 μL 100% ethanol and 7 μL 3 M NaAC, mixed and centrifuged for 10 min at 4 °C. The supernatant was further transferred into a solution containing 750 μL 100% ethanol and 43 μL 3M NaAC and stored at -20 °C for 30 min, followed by centrifugation for 10 min at 4 °C. The pellet was washed with 70% ethanol and air-dried. At last, 20 μL DEPC H_2O was added to dissolve the pellet. In this protocol, all the centrifugations were performed at 22,000 x *g*. The RNA samples were stored at -80 °C for further usage.

The same RNA isolation protocol was used for wheat and canola seeds except for the seed amount. For wheat, five seeds were frozen in liquid nitrogen and grounded into powder, and about 1/5 of the powder was used in the followed steps. Using five seeds might reduce the possible variation among individual seeds. For canola, ten seeds were frozen in liquid nitrogen and grounded into powder, and all the powder was used for RNA extraction.

2.4 Analysis of total RNA integrity

The quantity and purity for total RNA were measured using a NanoDrop 8000 Spectrophotometer (Thermo Fisher Scientific, http://www.thermofisher.com). The readings of 260/280, 260/230, and concentration were obtained for each RNA sample. While the maximal absorbance of nucleic acids is at 260 nm, the maximal absorbance of aromatic amino acids in protein is at 280 nm. The ratio of absorbance at 260 nm and 280 nm is frequently used for analyzing the purity of RNA, with "pure" RNA generally having a ratio of ~2.0. If there is contamination from protein, phenol or other components that absorb at 280 nm, the ratio would be lower. On the other hand, other impurities such as EDTA, carbohydrates and salts have absorbance at 230 nm and thus could decrease the ratio of 260/230.

Total RNA integrity was evaluated by agarose gel electrophoresis and Agilent 2100 Bioanalyzer (Agilent Technologies, http://www.agilent.com). In gel electrophoresis analysis, the relative brightness of 25S and 18S rRNA bands and the band smearing were considered as the major parameters. For high-quality total RNA, the brightness of 25S rRNA should be roughly twice that of 18S, while the bands should not show a heavy smearing. In the Bioanalyzer analysis, the RNA Integrity Number (RIN) is generated, based on the measurements such as the ratio of 25S and 18S rRNA, the relative area of 18S and 25S peak in the overall area, and the presence of unexpected peaks between the peaks for 18S and 25S rRNA (Schroeder *et al*., 2006).

2.5 Reverse transcription PCR (RT-PCR) and quantitative PCR (qPCR) analyses

The first-strand cDNA was synthesized from total RNA using the ThermoScript RT-PCR system and Oligo(dT) primer according to manufacturer's instructions (Invitrogen). Briefly, in a 0.5 mL PCR tube (with cap), 1.0 μg total RNA was mixed with 1 μL Oligo(dT) (50 μM), 2 μL dNTP Mix (10 mM), and DEPC-treated water to make a total mixture volume of 12 μL. The tube containing the mixture was incubated for 5 min at 65 °C. After adding a

mixture containing 4μL 5X cDNA Synthesis Buffer, 1μL 0.1 M DTT, 1μL RNaseOUT (40 units/μl), 1μL DEPC-treated water and 1μL ThermoScript reverse transcriptase (15 U/μl), the tube containing the 20 μL reaction mixture was incubated further under 50°C for 45 min. Following the incubation the synthesized cDNA was diluted 1:3 with sterile H_2O and used in polymerase chain reaction (PCR) amplification. The PCR products were analyzed by electrophoresis in 1% agarose gels.

The genes and primers used in qPCR analysis can be found in Table 2.2 (for Arabidopsis), Table 2.3 (for wheat) and Table 2.4 (for *Brassica napus*). qPCR was performed with the Green-2-Go qPCR Mastermix (BioBasic). The mastermix (12.5 μL for each), primers (in the final concentration of 0.25 μM each) and cDNA (1.5 μL for each) were added in the reaction volume of 25 μL each. Reactions were run in the Bio-Rad CFX96 Real-Time system and the threshold cycle (Ct) value for each reaction was generated by CFX Maestro Software (Version 3.0). Linear regression analysis was performed with Excel 2010 and ANOVA was conducted with RStudio (https://www.rstudio.com/).

Table 2.2 List of genes or fragments used in qPCR analysis on changes in stored mRNAs in aged Arabidopsis seeds.

Gene or fragment code	Gene ID	Sequence length (bps)	Forward primer	Reverse primer
A18	At1g74370	786	ATGTGGAATTTAGCTTCAAAAT	TCAGGTTGCAATCTCTCTG
A20	At3g01650	1470	ATGGGAGGAGGGAATTCCAA	TTAGTAAAGCTTGATTCTGGTC
A21	At5g14420	1407	ATGGGGACAGGGAATTCTAA	TTAGTAGAGCTTTATTCTTGTC
A26	At5g38895	981	ATGGGTGCTTTCTGTTGTTGC	CTAAGAAGTTTCATTTTCATCAAA
A27	At5g41350	639	ATGGGAGGTTGCTGTTGTT	CTAGTCAAGAGTAGAGTCAAAT
A31	At1g55530	1056	ATGGAAGAAGCAGTTGTAACG	TTAACTAGAATCCGTTGGTGC
A33	At2g17730	762	ATGGCCTCGTCATCATC	TTAAATATCACGTCTACACATC
A35	At5g47570	1400	ATGACGACGACGGCGGAA	CTAAAAGTGATGGTTCCCTC
A38	At4g00335	537	ATGGGAGGTTGCTGTAGTT	CTAGTTCAAGCGATCATCA
A40	At1g63170	1146	ATGTCAAGAGAGACGACGA	CTAGACTTCCTCGCCTTC
A41	At1g12760	1227	ATGTCAACGGAGACTACCAC	CTAGACTTCCTCTCGATCC
A42	At3g61180	1140	ATGAATCAATCGATCCCTTC	TCACACTTGTTCACTGCC
A44	At3G46510	1983	ATGGCCGGAGGAATCGTCT	TTATTGGCATGCTTTACGAAG
A45	At1G29340	2190	ATGGCCTCAGCGGCGATA	TCACAACACTGGTACGGAG
A47	At5G67530	1788	ATGGGGAAGAAACAACACAG	TTACCAGCTAGAGAAATCTTTA
B1	At4g10250	590	ATGATGAAGCACTTGCTAAG	TCAGAGTTCTTTGGATTCAG
B2	At5g59720	490	ATGTCTCTCATTCCAAGCAT	TCAATTAGCCCCGGAGATA
B4	At5g12030	461	ATGGATTTGGAGTTTGGAAG	TCAAGCGACTTGAACTTGTA
B5	At4g27670	685	ATGGCTTCTACACTCTCATT	CTACTGAATCTGGACATCG
B7	At1g16030	1940	ATGGCGACGAAATCAGAGA	CTAATCCACCTCTTCGATC
B9	At2g32120	1692	ATGGCAGAAGCAGCATACA	TTACATCTTTCTCTGAAGAGT
B10	At5g09590	2049	ATGGCTACCGCAGCTCTC	TCACTTTTTCACTTCCTCGT
B12	At5g56000	2100	ATGGCGGACGCAGAGACC	TTAGTCGACTTCCTCCATC
B13	At2g04030	2343	ATGGCTCCTGCTTTGAGTA	TCAATCTTGCCAAGGATCAC
B14	At3g07770	2400	ATGATCAGGCTCTCTAAGC	TCATTTCTTCCCATCCACTT
B16	At1g74310	2736	ATGAATCCAGAGAAATTCACA	TTAATCCTCGATCATTTCCT
B17	At5g15450	2907	ATGGCGACGGCTACGACGA	CTATTTTGAGAAAGCTGCTTC
B20	At5g51070	2838	ATGGAGGTGTTATCTACTTC	CTATGCGATCGATGTTTTGT
D5	At1g15690	2313	ATGGTGGCGCCTGCTTTG	TTAGAAGTACTTGAAAAGGATAC
$B16_{250}$	At1g74310	250	AGAACTGTCAAAGATGGTTGTG	TTAATCCTCGATCATTTCCTCA
$B16_{500}$		500	AGGTGAGGAAACACTTCAGAC	TTAATCCTCGATCATTTCCTCA
$B16_{1000}$		1000	TGCAGTTTCTGAGGCAATTCT	TTAATCCTCGATCATTTCCTCA
$B16_{1500}$		1500	ATTGATAACCTTGAAAGGAAGAG	TTAATCCTCGATCATTTCCTCA
$B16_{2000}$		2000	TAGTTGCTGGTGCTAAATACC	TTAATCCTCGATCATTTCCTCA
$B16_{2500}$		2500	ACCTGATGATATTCCAGCGAG	TTAATCCTCGATCATTTCCTCA
$B20_{250}$	At5g51070	250	TTTAGAGGTGTCTGAACCCGT	CTATGCGATCGATGTTTTGTC
$B20_{500}$		500	GATCATTGGCCATTGCAAAGG	CTATGCGATCGATGTTTTGTC
$B20_{1000}$		1000	TGAAAGAATGCTTCTTATGAGTC	CTATGCGATCGATGTTTTGTC
$B20_{1500}$		1500	TTTCAGCCAGTGTTGATTAACG	CTATGCGATCGATGTTTTGTC
$B20_{2000}$		2000	CACGTGCTAGTGAGGGTCTT	CTATGCGATCGATGTTTTGTC
$B14_{2000}$	At3g07770	2000	ATAAGCTTCGTTATTTGAGCGT	TCATTTCTTCCCATCCACTTC
$C22_{2000}$	At3g12280	2000	ATAGCTTACTTGGATCTGGAAG	CTATGAATCTGTTGGCTCGG
$B10_{2000}$	At5g09590	2000	CCTCTCCTTTCTCAGCTTAC	TCACTTTTTCACTTCCTCGTACT
$D5_{2000}$	At1g15690	2000	GAAAAGGATACCACCGTGAGT	GTGTCTTCATGATTTTCTTTGC

For the convenience of referring, each gene is given a simple code. The correlation between ΔCt value and fragment size was initially analyzed with a total of 29 Arabidopsis genes. The correlation was further confirmed with two genes B16 and B20. The fragments used for B16 are: $B16_{250bp}$, $B16_{500bp}$, $B16_{1000bp}$, $B16_{1500bp}$, $B16_{2000bp}$, and $B16_{2500bp}$, while the fragments for B20 are: $B20_{250bp}$, $B20_{500bp}$, $B20_{1000bp}$, $B20_{1500bp}$, and $B20_{2000bp}$. For comparisons of different genes, $B10_{2000bp}$, $B14_{2000bp}$, $B16_{2000bp}$, $B20_{2000bp}$, $C2_{22000bp}$, and $D5_{2000bp}$ were used.

Table 2.3 List of genes (or fragments) and primers used to analyze mRNA degradation in wheat seeds.

Gene code	Gene name	Coding Sequence Length	Fragment Analyzed	Forward primer	Reverse primer
W1	TraesCS7A02G070100	1815	950	TCAGGGAGCGGCGACGTTC	GCTACGACAAGCCGGTGGA
W2	TraesCS5D02G425100	1519	941	AAGTTGAACAAGTATGGTCGTC	GTATCTACCGGCTCGAACTC
W3	TraesCS2B02G309900	2785	950	CCTAGTGATGGAGTATTGTCC	CCACTAGAAGAGCTCGAACTC
W4	TraesCS4A02G296000	1767	949	GGGACTCACGGACGCAGAC	GGCGTTGATCTCGAAGCAC
W5	TraesCS2A02G027800	2274	951	CTCAGTTCATGTTCCCTGGT	TCACCACGTTGAGAAATGTCT
W6	TraesCS3A02G471900	2282	951	CGCTCGTCTGGTCCACGT	CGCTTCTTCTCCATGTCGAA
W7	TraesCS2B02G567600	678	500	CCTGCACCACCAGAAGCAC	TCCAGTTCACCACCTTCAG
W8	TraesCS1A02G045700	1479	950	CCATGCTCACCCTCCCGT	CGGCGGGAGGGCTTACGG
W9	TraesCS5B02G267900	2418	950	TCTGCAAAGGAGATTGATGAG	TTCATCTTCTTTGGACTACGTC
W10	TraesCS5B02G106300	2460	955	CTTCAAGAGAGTCGAATTGTGC	CAAGCGGCTCAGGTTCAGAGC
W11	TraesCS2D02G289100	2832	900	GCTGTGAAGAGATCACGCAC	TCAAAGCGTTGGTGTTGGATC
W12	TraesCS4A02G143200	2910	950	ATTATGCATTGAACCGAGGAGC	CTGGTCCTGGGTTGTCCTTC
W13	TraesCS7A02G517700	2070	2000	CCGTTTTCGCGGTCCTGATC	CTAAAGATACTTAAAGAGGATGC
W14	TraesCS2B02G521600	1458	900	GAGTTCAAGATCGTCCTCACC	GGATGCTAATGTAGTCAGACTG
W15	TraesCS7B02G068100	1605	948	GGCCTTATCTTTGACAAGAAGG	TTACCTTACAGTGACTATATCAT
W16	TraesCS4A02G100500	1128	950	GGTACATCTCCGCGGCGC	TGAACGAGCAATCTTCGCTGC
W17	TraesCS3B02G311900	1899	950	GACACTGCTGTAGAAGAGGC	TGCTCCGGTCTCTCGTTCC
W18	TraesCS3D02G321500	741	655	GCCTAGAGCGGCGGAGAAG	GGACAGATCAACGACCGACG
W19	TraesCS3A02G277700	2304	951	GATGGAGATGACCATTAAATAC	CTTGTCCTTGTCATCCGCAC
W2	TraesCS5D02G425100	1519	107	ATCTTGCTCGCGAAGGTAATG	GTATCTACCGGCTCGAACTC
			295	TGTTATACCGGTAGCTTCAGG	
			490	GGTATGCATTTCCGTGTATTAG	
			703	GCATTACTTGAATGCGACTGC	
			933	AAGTTGAACAAGTATGGTCGTC	
			1131	AGCCAATGGATCTGTTATGTAG	
W3	TraesCS2B02G309900	2785	950	CCTAGTGATGGAGTATTGTCC	CCACTAGAAGAGCTCGAACTC
W10	TraesCS5B02G106300	2460	955	CTTCAAGAGAGTCGAATTGTGC	CAAGCGGCTCAGGTTCAGAGC
W12	TraesCS2D02G289100	2910	950	ATTATGCATTGAACCGAGGAGC	CTGGTCCTGGGTTGTCCTTC

A total of 19 wheat genes were tested. For the convenience of referring, each gene is given a simple code. W2 was used to analyze the correlation between ΔCt value and fragment size, with the following fragments: $W2_{107bp}$, $W2_{295bp}$, $W2_{490bp}$, $W2_{703bp}$, $W2_{933bp}$, and $W2_{1131bp}$. Four genes with similar fragment lengths $W2_{950bp}$, $W3_{950bp}$, $W10_{955bp}$, and $W12_{950bp}$ were used.

Table 2.4 List of *B. napus* genes (or fragments) and primers used to study stored mRNA degradation during seed aging.

Gene code	Gene name	Coding sequence length	Fragment Analyzed	Forward primer	Reverse primer
Bn1	BnaA01g34230D	1659	1001	GGTCATAATAAGGTTTACAAGTC	TAGGTATCATATGAACAAGCTTG
Bn2	BnaC04g38000D	1080	1001	GGTCATAATAAGGTTTACAAGTC	TAGGTATCATATGAACAAGCTTG
Bn3	BnaC06g14690D	1062	1000	CTGTAACTGGATAACTAGTACTG	AGATGGAGCCTCAACAGCAG
Bn4	BnaC05g37990D	1470	1004	CGATCTGCACTATGATCCTTAC	GATTACGTGGTACAATTGCATC
Bn5	BnaA08g15530D	2301	1000	GAGTGCTTCTGCTTGATGGC	CAGTTGGTTAATATCTCAGCTC
Bn6	BnaC07g44190D	3145	1000	GTGCTGCAGAAGCGGAGAG	CATTCCAGAACCTCAAAGTCTC
Bn7	BnaC08g42010D	1493	1000	GAGATGATCCTTCCCATTCAC	CTTCAGCCATAATCCTCCTTG
Bn8	BnaC09g16410D	1443	928	GGTATAAGTTTATATGGTTCGAAC	TGGATTCGACCTTGTCACTTG
Bn9	BnaCnng08950D	1453	995	TATTCGTGGGTGGGAAAGGAG	TCAACTCTTGATCCGTTCCAG
Bn10	BnaA02g02620D	1965	1938	GGTGTATGGAGAATCTGCTGG	CAGCTGCTTATGCAACTCTTC
Bn11	BnaA06g10730D	1490	1485	ATGACTCTCAGAGACAGGCC	CAACCTCTTCTATCTTCGGAC
Bn12	BnaA09g55830D	2925	1851	AGCAGTACTGAAGGAAGTCAC	TCGATAGGAGCATTGTTCTCG
Bn13	BnaAnng00830D	1681	1486	CGTTACAACTCTGGTGCTGTC	AGAGCCTTGACGCATTGATTG
Bn14	BnaC01g16200D	2139	1488	ACCAAGGCTGTTATCACTGTG	GTCGATCACATCATCTCCATC
Bn15	BnaC01g39690D	2415	1920	CTATGCAGACAAGGAGAATGG	TCCACTTCAACTGGTTCAACG
Bn16	BnaC03g03860D	2046	1946	ATCACTGAACGGTCAGAACTG	TCTTCACTTCCTCGTAGTCAG
Bn17	BnaC03g13930D	1731	1731	AACAACACCTTGACCATCGTC	CATCTTGCTACCTTCAGCATC
Bn18	BnaC04g12620D	1692	1690	TGGCAGAACCAGCATACACAG	CACTTCCTCTGAAGAGTAACC
Bn19	BnaC06g23110D	2652	1800	CTATGTTAGCTAGAGGTCAGC	CCTCAATCCTCATCTTCTTCAC
Bn11	BnaA06g10730D	1490	1490	ATGACTCTCAGAGACAGGCC	CAACCTCTTCTATCTTCGGAC
			148	AGAGAGTGATCCAACAGATGG	
			487	ACTGTGTTGTGATTATGACCTC	
Bn12	BnaA09g55830D	2925	850	GTCTCTCTGATCCTAACCGTC	GTTGCTTGTCTGGTCAACATC
			1169	TCTGCTGAACGTGAATACGAC	
			1488	TACATCACTGAACGCTTCCTG	
Bn13	BnaAnng00830D	1681	1486	CGTTACAACTCTGGTGCTGTC	AGAGCCTTGACGCATTGATTG
Bn14	BnaC01g16200D	2139	1500	ACCAAGGCTGTTATCACTGTG	GTCGATCACATCATCTCCATC
Bn17	BnaC03g13930D	1731	1500	AACAACACCTTGACCATCGTC	CATCTTGCTACCTTCAGCATC

A total of 19 *B.napus* genes were used. For the convenience of referring, each gene is given a simple code. Five fragments on Bn12, $Bn12_{148bp}$, $Bn12_{487bp}$, $Bn12_{850bp}$, $Bn12_{1169bp}$, and $Bn12_{1500bp}$, were used to analyze the correlation of ΔCt value to the mRNA fragment size. For comparisons of different genes with similar fragment lengths, fragments $Bn11_{1490bp}$, $Bn12_{1500bp}$, $Bn13_{1486bp}$, $Bn14_{1500bp}$, and $Bn17_{1500bp}$ were used.

2.6 Estimating the relative amount of undamaged stored mRNAs during seed aging

Assuming that the qPCR amplification efficiency is nearly 100% for the early phase up to the Ct value (the threshold cycle reflecting the number of cycles required for the fluorescent signal to exceed the threshold background level) of qPCR, the amplified DNA copy number (designated as C here) can be estimated as:

$$C = N \cdot 2^{nc}$$

where N is the initial copy number of cDNA (cDNA is the proxy for undamaged mRNA or a selected region of mRNA), and *nc* is the number of PCR cycles used.

For the unaged control sample, the copy number of amplified DNA (C_c) for a specific fragment of mRNA at the threshold cycle can be estimated as:

$$C_c = N_c \cdot 2^{Ct(c)}, \quad (1)$$

Similarly for the aged sample, the copy number of amplified DNA (C_a) at the threshold cycle can be estimate as:

$$C_a = N_a \cdot 2^{Ct(a)} \quad (2)$$

Since the copy number of DNA at the threshold cycle can be considered the same for the control and aged samples ($C_a = C_c$), the relative amount of undamaged stored mRNA fragment in aged sample over unaged sample (i.e. N_a/N_c) can be estimated as:

$$N_a / N_c = 2^{Ct(c)} / 2^{Ct(a)}$$
$$= 2^{[Ct(c) - Ct(a)]}$$
$$= 1/2^{\Delta Ct(a-c)} \quad (3)$$

As the variance of this estimate is proportional to the variance of $\Delta Ct(a-c)$, the standard error of the estimate can be obtained from repeated measurements of the differences in their threshold cycles.

2.7 Estimating the rate of stored mRNA degradation at the one nucleotide level

We derived an estimator for the probability of stored mRNA degradation at the one nucleotide level. For simplicity, we assume that any change in mRNA to be equivalent to "a break on a nucleotide" if it prevents the mRNA template from being reverse-transcribed at the

particular nucleotide. Also, there are only two possible outcomes for an intact nucleotide after a given time of aging (say days): either it is broken at the probability β, or it is still intact with a probability of 1- β. Under these conditions, the probability for a stored mRNA with "*n*" number of nucleotides to have "*x*" number of break(s) would follow a binomial distribution and can be defined as:

$P(x) = C_x^n (t\beta)^x (1 - t\beta)^{n-x}$

where $P(x)$ is the possibility that x break(s) occur on the mRNA fragment,

n is the number of nucleotides in the fragment,

t is the aging time in days, and

β is the possibility that a break occurs on each nucleotide per day.

In reverse-transcription using oligo dT as the primer, only the mRNA templates without a break from the 3' end to the 5' target position could be reverse-transcribed into cDNAs and be further amplified in qPCR reaction. $P(0)$ can be defined as the probability that no nucleotide is broken within the given mRNA template, and thus we have:

$P(0) = C_0^n (t\beta)^0 (1 - t\beta)^{n-0}$

$P(0) = 1*1*(1 - t\beta)^n$

$P(0) = (1-t\beta)^n$ (4)

For a long fragment, it consists of a large number of nucleotides (with $n \rightarrow$ infinite), and the equation (4) can be further transformed into

$P(0) = \lim_{n\to\infty} (1 - t\beta)^n$

$$= \lim_{n\to\infty} \left(1 + \frac{1}{\frac{1}{-t\beta}}\right)^n$$

$$= \lim_{n\to\infty} \left(1 + \frac{1}{\frac{1}{-t\beta}}\right)^{\frac{1}{-t\beta} - t\beta n} \quad (5)$$

Since the expression of exponential growth constant e is $\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$, equation (5) thus can be simplified as:

$$P(0)= e^{-t\beta n} \quad (6)$$

Since the relative amount of undamaged stored mRNAs (with "0" break) can be estimated from equation (6), we have the following estimate of $P(0)$:

$$P(0)= e^{-t\beta n} \approx N_a/N_c = 1/2^{\Delta Ct} \quad (7)$$

Taking a logarithmic transformation of equation (7), we can estimate the probability (β) for one nucleotide to break in an mRNA of n-nucleotide long and aged for t time, as:

$$\beta = \ln(2^{\Delta Ct})/tn \quad (8)$$

2.8 Electrical conductivity (EC) Test

Ten mg Arabidopsis seeds were soaked into a 15 mL tube containing 1.5 mL ddH_2O. After 16 h incubation under room temperature (24 °C), the electrical conductivity (EC) was measured with VWR Bench/Portable Conductivity Meter (VWR, http://www.vwr.com). Three sets of data for each seed set were collected.

2.9 Analyses of seedling growth

Arabidopsis seeds were surface-sterilized with 20% bleach (8.25% sodium hypochlorite) and stored for 2 days at 4-5 °C in the dark for stratification. They were sown on medium containing ½-strength of Murashige & Skoog salts (Murashige and Skoog, 1962), 1% w/v sucrose and 0.7% w/v agar. The plates were placed in a growth chamber (16h/8h photoperiod and light intensity of 90 ± 10 μM m^{-2} sec^{-1}) at 20 °C. After 7 days, a seed was considered germinated if the radicle was equal or longer than the length of the seed. For the fresh seedling weight, all seedlings from one plate were weighed together and the average seedling weight was calculated. For the root length, seeds were sown on the plates, which were placed vertically in the growth chamber. After 10 days, plate images were taken, the primary root length of each seedling was measured with the NIH ImageJ software (Version 1.42) and the average root length per plate was calculated. At least three plates were used for each seed sample.

2.10 MinION nanopore sequencing

2.10.1 cDNA synthesis, template enrichment, and purification

The major steps in the cDNA-PCR sequencing (SQK-PCS108) protocol for MinION sequencing (https://community.nanoporetech.com/protocols) were followed with some modifications. Reverse transcription was performed with the template-switching SuperScript II Reverse Transcriptase (RTase; Invitrogen). The reaction was initiated with a mix of 1.0 μg total RNA, 1.0 μL dNTP (10 mM), 1.0 μL poly(T18)-VN primer (5 μM), and ddH_2O up to a total volume of 10.5 μl. The mixture was incubated at 72 °C for 2 min, and then cooled down to 47 °C, followed by the addition of a mixture of 4.0 μL of the 5x first strand buffer, 2.0 μL DTT (20 mM), 0.5 μL RNase OUT and 1 μL SuperScript II RTase. After 50 min incubation, a mixture consisting of 0.5 μL template-switching primer (5 μM) and 0.5 μL SuperScript II RTase was added, followed by incubation at 42 °C for 60 min and heat inactivation at 70 °C for 10 min. The cDNA was enriched by 15 cycles of PCR (94°C: 30 sec; 60:1 min; 72°C: 3 min) using primers specific for the oligo dT primer and an adaptor sequence on the template switching primer. The enriched cDNA was then purified using KAPA Pure Beads (Roche; Kit code: KK8002). For the purification, 150 μL KAPA Pure Beads were mixed with 50 μL PCR products. After a brief vortex and 15 min incubation on a rotator, the KAPA Beads were concentrated with a magnet plate. After removal of the supernatant, 80% ethanol (200 μl) was added into the tube containing the KAPA Bead pellet. After 30 sec incubation on the magnet plate, ethanol was removed, and the pellet was washed once again with 80% ethanol. Then the pellet was dried and re-suspended with 17 μL water. After 5 min incubation off the magnet plate, the tube was placed on the magnet plate to concentrate the KAPA beads. The supernatant containing purified cDNA templates was then transferred to a new tube. The concentration was determined by a NanoDrop spectrophotometer.

2.10.2 cDNA end-repair

The cDNA prepared above (1 μg) was added with nuclease-free water to a final

volume of 45 µL and further prepared for MinION sequencing with NEB End repair/dA tailing kit (New England Biolabs). A 60 µL mixture was prepared in a 0.2 mL PCR tube, consisted of the 45 µL purified cDNA, 7 µL Ultra II end-prep reaction buffer, 5 µL DNA control strand (Oxford Nanopore Technologies), and 3 µL Ultra II end-prep enzyme mix. The mixture was incubated at 20 °C for 5 min and 65 °C for 5 min. The end-repaired product was purified using 180 µL KAPA Pure Beads as described above, with two washes of 80% ethanol. The DNA was eluted in 36 µL ddH_2O with 5 min incubation at room temperature.

2.10.3 Adapter ligation to cDNA ends

The adapters were ligated to cDNA ends in a mixture containing 30 µL end-repaired DNA 20 µL adaptor mix (Oxford Nanopore Technologies), and 50 µL Blunt/TA ligation Master Mix (New England Biolabs) through a 10 min incubation at room temperature. The DNA was purified using KAPA beads (300 µl). The KAPA beat pellet was washed once with Adapter Bead Binding buffer (Oxford Nanopore Technologies), and DNA was eluted by the addition of 15 µL Elution Buffer (Oxford Nanopore Technologies) was added and then incubated for 15 min at room temperature.

2.10.4 Preparing preloading mixture and priming flow cell

The pre-loading mixture was prepared, containing 11 µL of eluted cDNA from the last step, 35 µL Running Buffer with Fuel mix (RBF, Oxford Nanopore Technologies), 25.5 µL Library loading Beads (Oxford Nanopore Technologies), and 2.5 µL ddH_2O. Before loading the pre-loading mixture into MinION SpotON Flow Cell MK I (R9.4; Oxford Nanopore Technologies), the flow cell was primed initially with 800 µL pre-prepared priming mixture (480 µL RBF and 520 µL ddH_2O) for 5 min and then with another 200 µL priming mixture. Finally, 75 µL pre-loading mixture was loaded into the flow cell. The flow cell was connected to a computer.

2.10.5 Acquisition, processing and analysis of the sequencing reads

The sequencing was performed with the sequencing app (MinKNOW; Oxford

Nanopore Technologies), using the sequencing program "NC_48Hr_Sequencing_Run_FLOMIN106_SQK-LSK108" in MinKNOW 1.4.2. After raw data collection, base-calling was performed with Albacore 0.8.4 (Oxford Nanopore Technologies). Further, sequencing reads were aligned to Arabidopsis transcriptome (TAIR10_cdna_20101214_updated, http://www.arabidopsis.org) using the Burrows–Wheeler Aligner 0.7.15 (http://bio-bwa.sourceforge.net/, released 6/1/2016). For data processing, the transformation of the data format was conducted with SAMtools (1.4, http://www.htslib.org/, released 13 March 2017), while the information extraction and manipulation of raw sequencing reads were conducted with Perl scripts (Supplementary file 1). Integrative Genomics Viewer (IGV_Win_2.3.92) was used to view the sequencing reads. Sequencing with MinION Flow Cell was conducted under Win 7, while other data manipulation including basecalling, format transformation, alignment and statistical analyses were performed under BioLinux 8 (Field *et al*., 2006).

3. CHAPTER THREE – SEED STORED mRNA DEGRADATION IN ARABIDOPSIS SEEDS

3.1 Introduction

As discussed in Chapter 1, previous research has provided some evidence to show that stored mRNAs are important for early stages of seed germination. Since mRNAs are prone to degradation in general, we postulated that seed stored mRNAs would undergo degradation during seed aging. Thus, one main objective of my thesis research is to determine how stored mRNAs change during seed aging process. First, I would use *Arabidopsis thaliana*, the most widely used model for studies of plant biology. Then, I will expand the studies to two crop plants, wheat (*T. aestivum*) and canola (*B. napus*), with a much larger genome than *Arabidopsis thaliana*.

When seeds are stored at 4-5 ºC in a refrigerator, they usually take a long time (typically years) to lose the ability to germinate completely, and even longer if stored at -20 ºC especially if the seed water content is low enough (Walters, 2008). Thus, for seed aging studies, accelerated (or artificial) aging treatments are often used so that seeds could age in a much shorter time. Also, the same seed resource can be used for AA seed treatments, reducing variation due to different seed batches. However, there is a question whether the changes in seeds under accelerated aging conditions truly reflect the changes under naturally aging conditions. In this project, we will use AA seeds and at the same time, Arabidopsis seeds were collected and stored in a refrigerator (cooler; 4-5 ºC) for various length of times, providing an opportunity to compare NA and AA seeds. Also for this work, having RNA samples of high quality from dry seeds is important. However, it is known that RNA isolation from dry seeds is more challenging than RNA isolation from other tissues such as leaves. Therefore, at the initial stage of the project, I compared different methods for RNA isolation from Arabidopsis dry seeds and selected one method for further improvements. In the end, we have a method which can produce RNA samples of high quality from unaged and aged seeds.

With a good RNA isolation method developed, and NA and AA seeds available, I

investigated the changes of stored mRNAs during seed aging as they relate to seed aging time and seed germination. New methods (mathematical equations) were developed for quantifying the changes in stored mRNA levels and the rate of changes at the one nucleotide level. My results show that Arabidopsis seed stored mRNAs are degraded at a constant rate under the same aging conditions, while the rate of degradation is highly dependent on the aging conditions such as the temperature. Further, the change in one stored mRNA can be used as a good measurement of seed aging.

3.2 Results

3.2.1 Seed aging treatments

3.2.1.1 Accelerated aging assay

Accelerated seed aging was performed similar to what was described (Sugliani *et al*., 2009), using high temperature (37 ºC) and high humidity conditions (about 83% relative humidity) (for details, see Materials and Methods). In the initial experiment, seeds were treated for 5, 10, 15 and 20 days (time points were chosen based on previous work done by a former graduate student, Shefeng Qian, in Dr. Wang's lab). For each time point, three tubes with seeds were treated. Following the treatment, 300 seeds from each tube were sterilized and plated onto the ½ MS plates (100 seeds per plate) for the germination test. The germination results are presented in Figure 3.1. The germination percentage decreased from nearly 100% to 92.8%, 37%, and 1.3%, after 5, 10, and 15 days of incubation, respectively. No seeds germinated after 20 days of AA treatment. The speed of decrease in germination percentage in this AA assay was similar to what was observed previously, demonstrating good reliability of the assay. Since seed aging could be controlled more precisely in terms of the seed quality and aging conditions, and the aging time is much shorter compared to the natural aging under normal storage conditions, AA Arabidopsis were used mostly in this study.

3.2.1.2 Seed aging under normal storage conditions at low temperatures

Arabidopsis seeds have been collected in Dr. Wang's lab over the years and stored in a

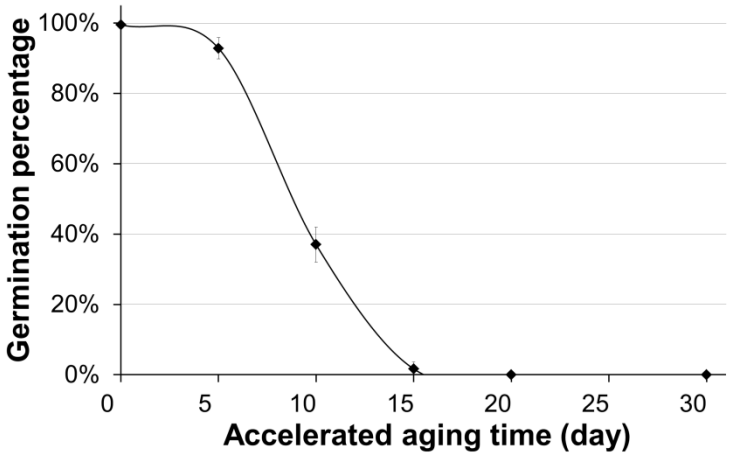


Figure 3.1 Effect of accelerated aging conditions on germination percentage of freshly harvested Arabidopsis seeds. Dry Arabidopsis seeds were acceleratedly aged for 0, 5, 10, 15, 20 and 30 days. For each seed lot, seed germination test was conducted with three replicate plates with 100 seeds each and the standard errors are shown.

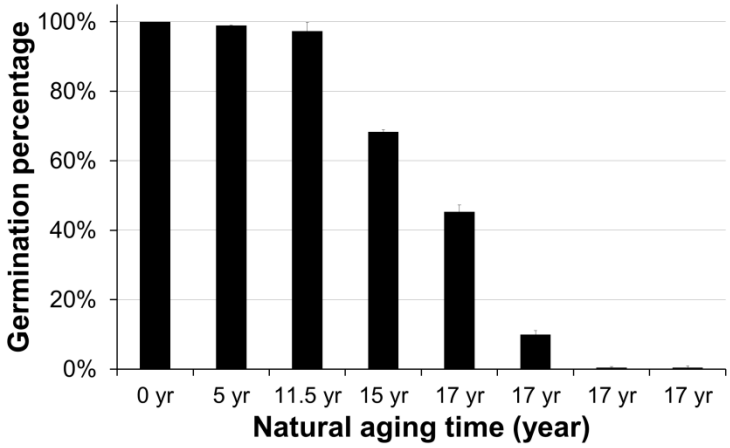


Figure 3.2 Effect of natural storage on seed germination percentage. Arabidopsis seeds have been collected in Dr. Wang's lab over the years and stored in a 4 ℃ refrigerator. Totally eight seed lots were used and their aging years were shown. The four seed lots of 17 years are (from left to right) 51, 26-14, 28-16, and 29-17 respectively (Table 2.1). Although they were all stored for 17 years, minor differences in seed quality or water content might exist at harvest since the plant growth and seed harvest were not strictly controlled to be the same for the different seed lots as they were collected initially for other purposes. A small difference in seed quality or water content can affect seed aging speed resulting in a difference in seed germination percentage. For each seed lot, seed germination test was conducted with three replicate plates with 100 seeds each and the standard errors are shown.

4 ℃ refrigerator (Table 2.1). As presented in Figure 3.2, the germination percentage decreased from nearly 100% for the recently harvested seeds to 0% for a 17-year seed stock. Apparently, seed aging under 4-5 ℃ and dry conditions is a very slow process. Also, different seed lots may differ in seed quality and resistance to the aging process. As shown in Figure 3.2, four seed samples which had been stored for 17 years in the same refrigerator had the germination percentages from 40% to 0.5%. The availability of AA and NA seeds with a range of seed germination percentage provided a good opportunity to compare these two different aging types.

3.2.2 Comparison of methods for RNA isolation from Arabidopsis seeds

Three protocols were compared initially to identify an appropriate one for isolating total RNA with good quality and acceptable quantity from dry Arabidopsis seeds. For this project, RNA integrity is fundamental to the subsequent analysis on mRNA degradation. However, RNA isolation from dry plant seeds is known to be more difficult than from fresh plant tissues such as leaves and flowers. To find a suitable protocol, three different methods were compared first: the protocols developed by Onate-Sanchez and Vicente-Carbajosa (2008), by Mornkham *et al*. (2013), and Qiagen RNAeasy Plant Mini Kit (catalog #74904; following the manufacturer's instructions). For convenience, the three methods are referred to as O.S, MMY, and QIAGEN method respectively. As shown in Figure 3.3 and Table 3.1, the O.S method produced better results than the MMY method and the Qiagen protocol in terms of the RNA quality and yield.

3.2.3 Improvement of an RNA isolation method

The O.S method was thus selected for further improvements. Every step of the protocol was tested and as a result several modifications were made. For convenience, the major steps in the original protocol (Onate-Sanchez and Vicente-Carbajosa, 2008) are described as follows:

1. The seeds were frozen with liquid nitrogen and ground with mortar and pestle, followed by a powder transfer to a cooled standard microfuge tube.

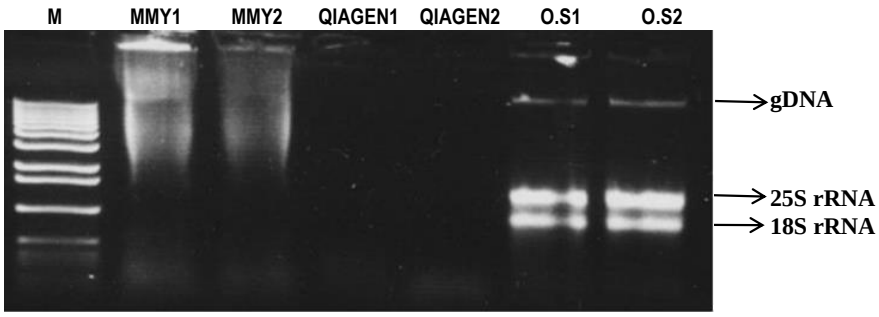


Figure 3.3 Comparison of three protocols for RNA isolation from dry Arabidopsis seeds. Total RNAs were isolated from dry Arabidopsis seeds using three protocols; they are MMY (Mornkham *et al*., 2013), QIAGEN (Qiagen RNAeasy minprep kit), and O.S (Onate-Sanchez and Vicente-Carbajosa, 2008) method respectively. For each method, two repeats were conducted; the six RNA samples from the same seed resource are referred to as MMY1, MMY2, QIAGEN1, QIAGEN2, O.S1, and O.S2 respectively. The RNAs were subjected to electrophoresis and visualized in a 1.0% agarose gel.

Table 3.1 RNA quality test for the total RNAs isolated from three methods.

Sample	260/280	260/230	Conc. (ng/ µ L)
MMY1	1.61	0.84	487.00
MMY2	1.49	0.49	919.00
QIAGEN1	1.55	0.14	6.10
QIAGEN2	2.45	0.03	15.20
O.S1	1.72	1.16	141.80
O.S2	1.78	1.47	131.10

Total RNAs were isolated from the same source of Arabidopsis seeds using three different methods (they are MMY, QIAGEN, and O.S method). For each method, two repeat RNA samples were obtained. The RNA quality and concentration were analyzed using a NanoDrop 8000 spectrophotometer.

2. Add 550 µL of chloroform and extraction buffer containing 0.4 M LiCL, 0.2 M Tris (pH 8.0), 25mM EDTA and 1% SDS; vortex the mixture 10 sec and centrifuge for 3 min.

3. The supernatant was transferred to a new 1.5 mL tube, and 500 µL of water-saturated acidic phenol was added and vortex thoroughly. After adding 200 µL chloroform, the mixture was centrifuged for 3 min.

4. The supernatant was transferred to a new 1.5 mL tube, and 1/3 volume of 8 M LiCl was added and mixed well. After 1 h incubation at -20 °C, the mixture was centrifuged for

30 min at 4 °C.

5. The pellet was treated with DNase I (10 units) for 30 min at 37 °C.

6. A mixture of 470 μL DEPC-H_2O, 7 μL 3 M NaAc (pH 5.2) and 250 μL ethanol was added into the DNase treated RNA sample. The tube was then centrifuged for 10 min (22,000 x *g*) at 4 °C to precipitate carbohydrates.

7. The supernatant was transferred to a new tube, and a mixture of 750 μL ethanol and 43 mL 3M NaAc (pH 5.2) was added. After 1 h incubation at -20 °C, the tube was centrifuged for 20 min centrifuge at 4 °C, and pellet was washed with 70% ethanol and air-dried. The RNA pellet was re-suspended with 20 μL DEPC-H_2O.

After several rounds of testing, the following modifications were made:

1. Washing with 70% ethanol was added after Step 4. This additional washing step helped to increase RNA purity and the subsequent DNase digest.

2. Shorter times were used in several steps, including step 4 (from 1 h to 30 min) and step 7 (from 1 h to 30 min), DNA digestion in step 5 (from 30 min to 20 min), and centrifugation times in step 4 (from 30 min to 15 min) and step 7 (from 20 min to 15 min).

With these modifications, total RNA samples of high and consistent quality were isolated from unaged, AA, and NA dry Arabidopsis seeds (Figure 3.4 and Table 3.2).

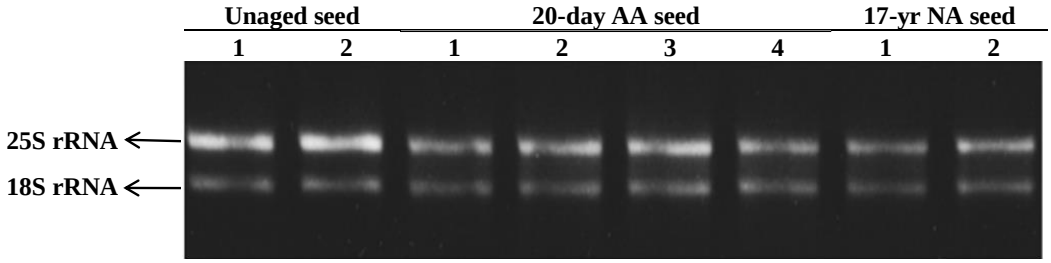


Figure 3.4 Example of total RNAs isolated from unaged and aged Arabidopsis dry seeds. Total RNAs were isolated from Arabidopsis dry seeds using the modified protocol as described in the method. Three types of seeds were used; they were unaged seeds, 20-day AA seeds, and 17-yr NA seeds and there were two total RNA samples from unaged seeds, four from 20-day AA seeds, and two from 17-yr NA seeds. The same amount of total RNA (0.5 μg) was loaded into each lane and subjected to electrophoresis.

Table 3.2 RNA quality test for the total RNAs isolated from unaged, AA and NA seeds using the modified method.

Seed sample	260/280	260/230	Conc. (ng/ μ L)
Unaged 1	2.04	1.85	116.00
Unaged 2	2.04	1.70	119.00
20-day AA 1	2.03	1.81	189.00
20-day AA 2	2.00	1.93	209.10
20-day AA 3	2.08	1.94	331.50
20-day AA 4	2.12	1.98	238.00
17-yr NA	1.94	1.97	137.30
17-yr NA	1.95	1.85	155.00

Total RNAs were isolated from unaged, AA, and NA seeds using the modified O.S protocol. For the three types of seed lots, two, four, and two repeats were conducted and the samples were referred to as Unaged1, Unaged2, etc. The RNA quality was tested using a NanoDrop 8000 spectrophotometer.

The RNA quality was also evaluated using Agilent Bioanalyzer since it can assess RNA integrity based on the RNA integrity number (RIN) and it has been used by others to evaluate RNA degradation in plant seeds (Fleming *et al*, 2018a; Fleming *et al* 2016). The total RNA samples from seeds with different seed germination percentages were analyzed with Agilent Bioanalyzer. The results indicated that the RNA integrity numbers (RINs) were very similar (mostly between 8 – 9) among RNA samples from seeds with different germination percentage (Table 3.3). These results indicate that revised RNA isolation protocol as described above produced RNA samples with consistent quality.

Table 3.3 RNA integrity numbers (RINs) of RNAs isolated from unaged and aged Arabidopsis seeds.

Germination Percentage (%)		NA							AA					
		100	99.0	87.3	68.3	46.3	10	0	100	92.7	52.0	20.7	3.7	0
25S/18S ratio	Repeat 1	1.73	1.75	1.68	1.58	2.63	1.85	1.66	1.81	1.62	1.83	2.14	1.63	1.79
	Repeat 2	1.67	1.71	1.5	1.25	2.05	1.5	2.08	1.54	1.7	1.74	1.45	1.78	1.91
RIN	Repeat 1	8.3	8.8	8.5	8.5	8.4	8.8	7.2	9.4	8.5	8.5	8.6	8.6	8.7
	Repeat 2	8.3	9.2	8.2	8	8.4	8	8.2	8.2	8.3	8.3	8.1	8.6	8.8

Total RNAs were isolated from naturally aged (six NA seed lots with the indicated germination percentages) and acceleratedly aged (five AA seed lots with the indicated germination percentages) as well as unaged control Arabidopsis seeds. For each aging-time point, usually two different RNA samples were used for RNA analysis by Agilent 2100 Bioanalyzer to obtain the 25S/18S ratio and RIN.

3.2.4 RT-PCR analysis of seed stored mRNA levels in aged Arabidopsis seeds

3.2.4.1 Survey of genes for the presence of stored mRNAs in dry seeds

Genomic analysis revealed more than 12,000 stored mRNA species in Arabidopsis dry seeds (Nakabayashi *et al*., 2005). Thus, many genes have stored mRNAs in mature seeds. However, it is understandable that these stored mRNAs may be present at different levels. Thus, to determine which genes have stored mRNAs at a level that can be readily detected by PCR for further analyses, we surveyed 120 genes which were used previously in our other research activities. Those genes were loosely grouped into four groups, they are the E2/E3 genes involved in protein ubiquitination, heat shock protein genes, cell cycle and growth related genes, and other genes, referred to as groups A, B, C and D respectively. For ease of identification, genes were referred to by the simple codes such as A5, B10, etc (Table 3.4). Since they belong to many different gene families and have diverse sequences, changes in stored mRNAs of these genes should reflect the changes occurring to stored mRNAs in general. Results from the initial survey indicated that stored mRNAs could be detected for more than half of the genes surveyed (Figure 3.5).

Table 3.4 List of genes used to survey the presence of seed stored mRNAs by RT-PCR.

E2/3(group A)			Heat shock proteins (group B)			Cell cycle and growth (group C)			Others (group D)		
Gene code	Gene ID	cDNA (bps)	Gene code	Gene ID	cDNA (bps)	Gene code	Gene ID	cDNA (bps)	Gene code	Gene ID	cDNA (bps)
A1	At1g78870	441	B1	At4g10250	590	C1	At3g48750	885	D1	At3g62060	1260
A2	At1g16890	489	B2	At5g59720	490	C2	At5g10270	1518	D2	At2g46930	1251
A3	At1g23260	477	B3	At1g59860	460	C3	At1g73690	1197	D3	At3g16090	1479
A4	At2g36060	441	B4	At5g12030	461	C4	At1g44110	1383	D4	At5g27150	1617
A5	At5g41700	450	B5	At4g27670	685	C5	At1g77390	1329	D5	At1g15690	2313
A6	At5g56150	447	B6	At5g02490	1962	C6	At5g43080	1068	D6	At2g01980	3441
A7	At1g17280	714	B7	At1g16030	1940	C7	At2g17620	1290	D7	At2g39800	2154
A8	At5g62540	453	B8	At4g24280	2160	C8	At1g16330	1947	D8	At1g28120	921
A9	At5g59300	612	B9	At2g32120	1692	C9	At1g70210	1020	D9	At1g18860	1443
A10	At5g05080	756	B10	At5g09590	2049	C10	At2g22490	1089	D10	At1g13930	468
A11	At1g36340	465	B11	At5g52640	2118	C11	At4g34160	1131	D11	At2g39730	1425
A12	At3g24515	1230	B12	At5g56000	2100	C12	At5g65420	957	D12	At4g20360	1431
A13	At5g41340	561	B13	At2g04030	2343	C13	At4g37630	972	D13	At2g30870	648
A14	At1g63800	558	B14	At3g07770	2400	C14	At4g03270	909	D14	At3g26650	1191
A15	At2g46030	552	B15	At4g24190	2472	C15	At2g23430	576	D15	At3g26650	1737
A16	At5g33393	1308	B16	At1g74310	2736	C16	At3g50630	630	D16	At2g43030	816
A17	At5g33393	837	B17	At5g15450	2907	C17	At3g24810	570	D17	At5g11670	1767
A18	At1g74370	786	B18	At2g25140	2895	C18	At3g19150	591	D18	At1g54870	1008
A19	At4g14220	1116	B19	At3g48870	2859	C19	At1g49620	588	D19	At5g03860	1689
A20	At3g01650	1470	B20	At5g51070	2838	C20	At5g48820	723	D20	At3g08030	1098
A21	At5g14420	1407				C21	At2g32710	870			
A22	At2g42360	711				C22	At3g12280	3042			
A23	At2g47700	1077				C23	At1g59580	1131			
A24	At4g11680	1173				C24	At1g07370	1025			
A25	At5g37270	627				C25	At2g29570	795			
A26	At5g38895	981				C26	At1g04820	1353			
A27	At5g41350	639				C27	At5g23860	1350			
A28	At1g73760	1104				C28	At3g16640	507			
A29	At3g58720	801				C29	At1g19850	2709			
A30	At5g08139	1131				C30	At1g30330	2808			
A31	At1g55530	1056				C31	At1g59750	1998			
A32	At3g20395	672				C32	At1g19220	3261			
A33	At2g17730	762									
A34	At5g73270	627									
A35	At5g47570	378									
A36	At5g05130	2589									
A37	At4g23450	627									
A38	At4g00335	573									
A39	At3g16090	1479									
A40	At1g63170	1146									
A41	At1g12760	1227									
A42	At3g61180	1140									
A43	At1G23030	1839									
A44	At3G46510	1983									
A45	At1G29340	2190									
A46	At2g35930	1236									
A47	At5G67530	1788									
A48	At5G61560	2391									

A total of 120 genes were tested and loosely grouped into four groups: 48 genes in group A, 20 genes in group B, 32 genes in group C, and 20 other genes in the last group (group D). For the convenience of referring in this study, each gene was given a gene code. The gene ID and length of cDNA coding sequence (CDS) in base pairs (bps) are indicated.

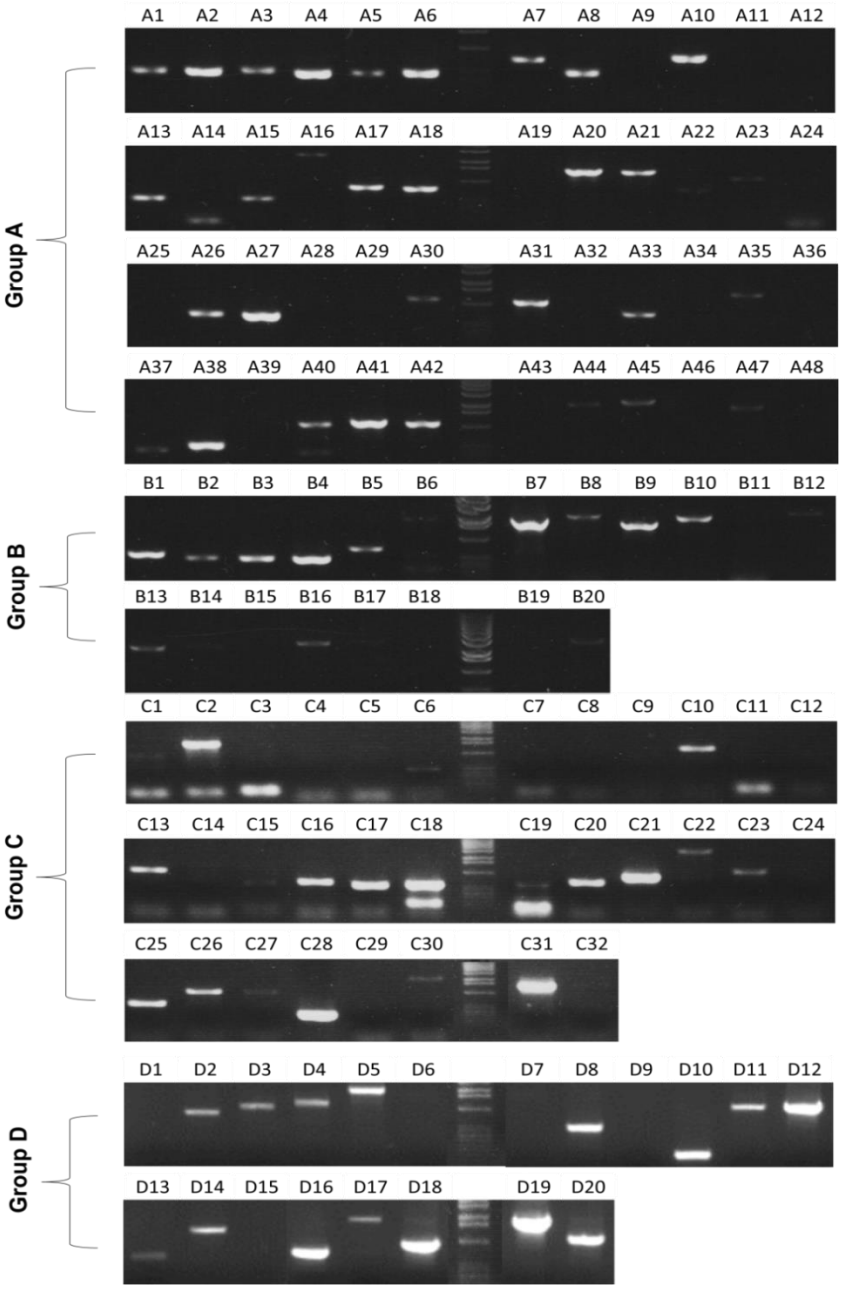
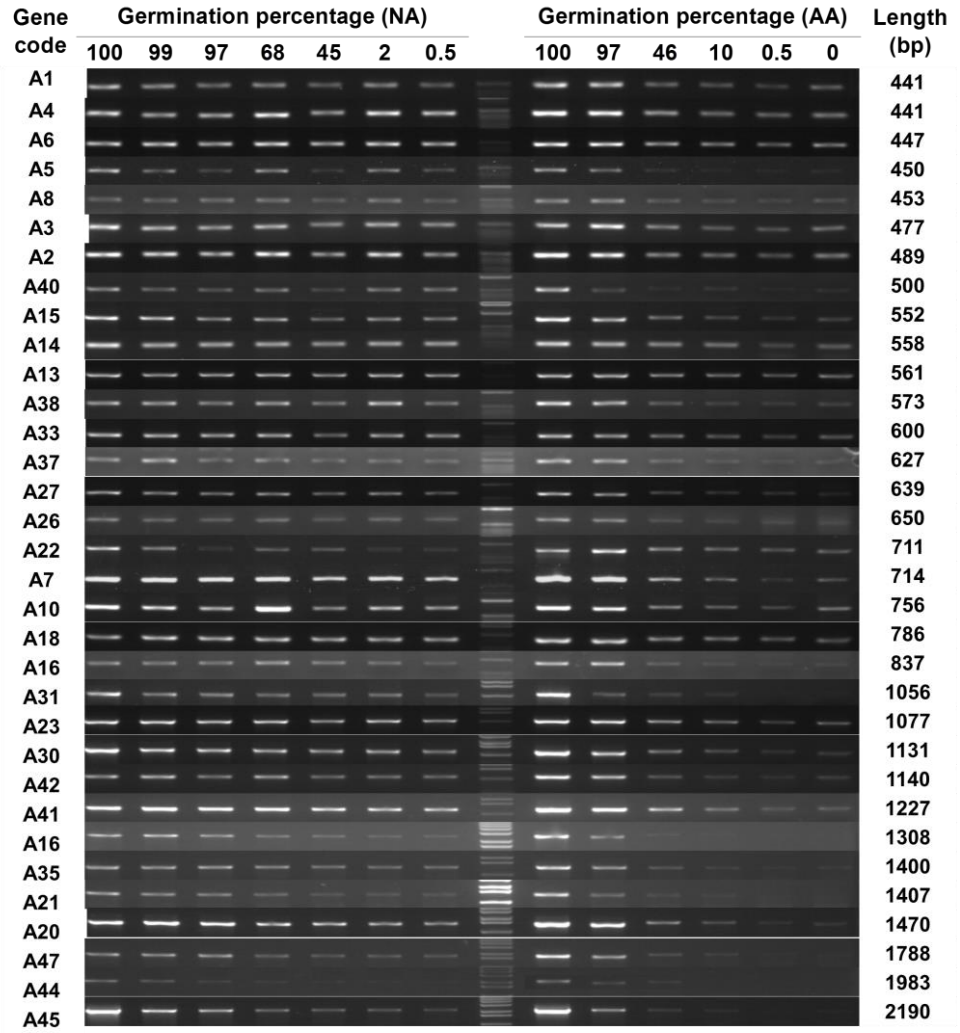


Figure 3.5 Results of RT-PCR analysis of 120 genes (in A, B, C and D groups) to survey the presence of stored mRNAs in Arabidopsis dry seeds. Information on these genes is provided in Table 3.4. cDNA reverse-transcribed from total RNA of unaged dry seeds was used in PCR with primers specific for each of the genes. The PCR products were subjected to electrophoresis in 1% agarose gels.

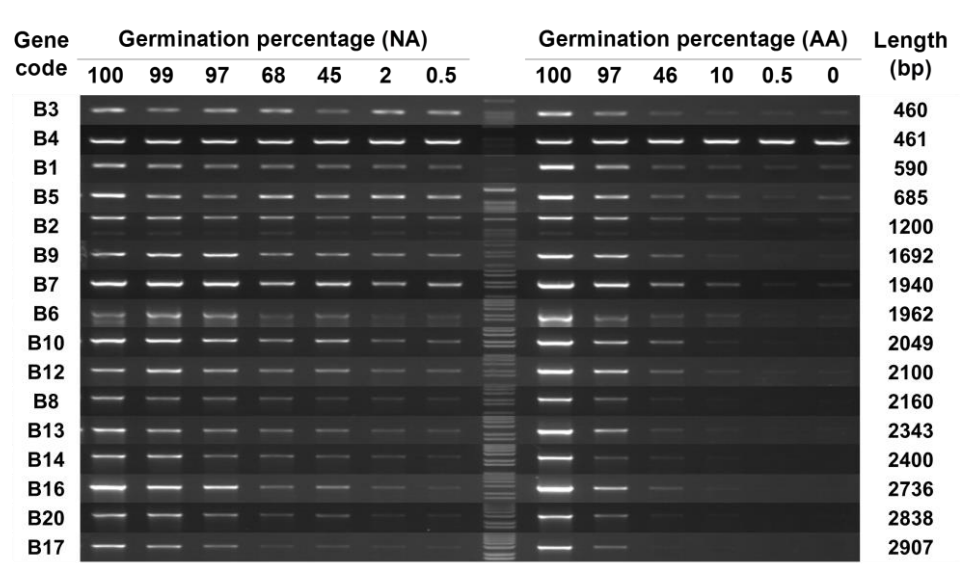
3.2.4.2 RT-PCR analysis of seed stored mRNA levels during seed aging

A number of genes (33, 16, 19 and 15 genes from A, B, C and D groups, respectively) were then used to determine the changes in stored mRNAs during seed aging. Since cDNAs were synthesized from stored mRNAs, the damage or lesion occurring to a stored mRNA during seed aging is reflected by the change in its cDNA level. cDNAs from two sets of seeds, NA and AA seed lots with different germination percentages as well as unaged control seeds, were used. They were amplified by PCR and the products were visualized following electrophoresis. The RT-PCR results showed that almost all stored mRNAs showed a gradual decrease in both NA and AA seed samples, but the extent of decrease varied greatly among different genes (Figure 3.6). With further analysis, it became clear that longer cDNA fragments showed a greater decrease in general. For easy viewing, genes in Figure 3.6 were presented according to the length of cDNA coding region used in the analysis. The trend was more apparent for the genes in group B, with the mRNA coding region ranging from 460 bp to 2900 bp (bp stands for base pairs) while most of the genes in group A have a length less than 1500 bp. Thus, the stored mRNAs showed a decrease during both natural and accelerated seed aging process, with longer mRNAs showing a greater decrease. At the same time, some cDNAs appeared not to show a decrease in aged seeds. However, a possible difference could be masked by PCR over-amplification or some variation in RT-PCR, which is only a "semi-quantitative" method. Nevertheless, the RT-PCR results provide a good overview regarding the changes in stored mRNAs during seed aging. Collectively, the RT-PCR results of various genes suggest that (1) most stored mRNAs showed a decrease during both natural aging and accelerated aging processes, and (2) longer mRNAs were more likely to show a greater decrease.

A


Gene code
Germination percentage (NA)
100 99 97 68 45 2 0.5
Germination percentage (AA)
100 97 46 10 0.5 0
Length (bp)
A1 441
A4 441
A6 447
A5 450
A8 453
A3 477
A2 489
A40 500
A15 552
A14 558
A13 561
A38 573
A33 600
A37 627
A27 639
A26 650
A22 711
A7 714
A10 756
A18 786
A16 837
A31 1056
A23 1077
A30 1131
A42 1140
A41 1227
A16 1308
A35 1400
A21 1407
A20 1470
A47 1788
A44 1983
A45 2190

B


Gene code
Germination percentage (NA)
100 99 97 68 45 2 0.5
Germination percentage (AA)
100 97 46 10 0.5 0
Length (bp)
B3 460
B4 461
B1 590
B5 685
B2 1200
B9 1692
B7 1940
B6 1962
B10 2049
B12 2100
B8 2160
B13 2343
B14 2400
B16 2736
B20 2838
B17 2907

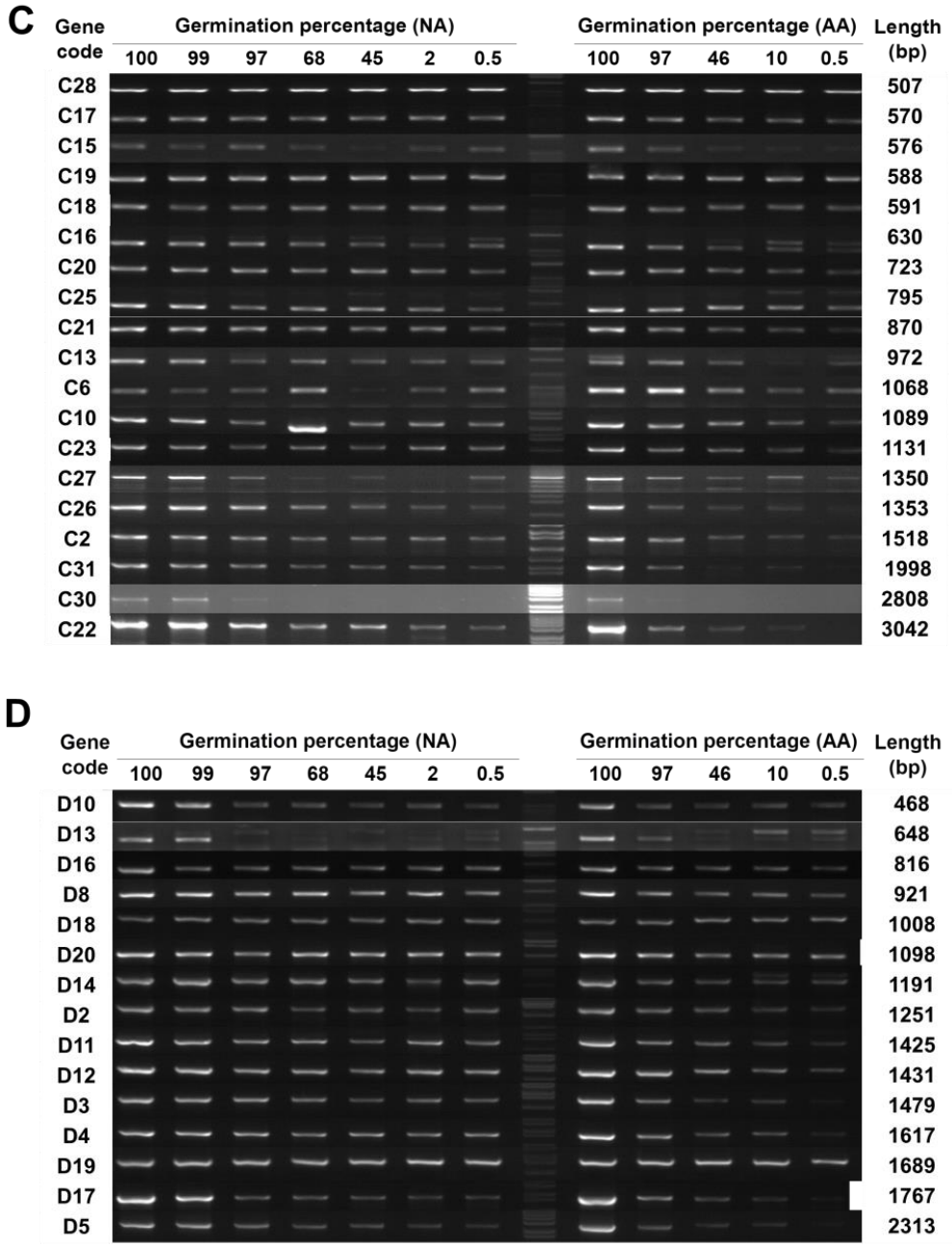


Figure 3.6 The level of stored mRNAs in naturally aged (NA) and acceleratedly aged (AA) Arabidopsis seeds detected by RT-PCR. cDNAs from NA (six NA seed lots) and AA (four AA seed lots) seeds with different percentages of germination as well as the unaged control seeds (100% germination) were used. Different genes in the two groups were analyzed. The number of PCR cycles varied depending on their mRNA abundance in the unaged seeds. PCR products were electrophoresed in agarose gels. (A) Analysis of 33 genes in group A (see Table 3.4 for the list of genes). (B) Analysis of 16 genes in group B. (C) Analysis of 19 genes in group C (for the list of genes, see Table 3.4). (D) Analysis of 15 genes in group D. The germination percentage of a seed sample is indicated above the DNA lane. M stands for the 1 kb DNA molecular ladder. The gene code is indicated at the left and the cDNA length amplified by PCR indicated at the right of each row.

3.2.5 qPCR analysis of seed stored mRNA levels in aged seeds

To quantitatively analyze the changes in stored mRNAs during seed aging, quantitative real-time PCR (qPCR) was used. Real-time detection of PCR products is made possible by including fluorescent molecules in the reaction, which bind to double-stranded (dsDNA). At the beginning of PCR reaction, the dsDNA products are not enough and thus the florescence cannot be detected by the qPCR system. As PCR reaction proceeds, products accumulate and finally yield a detectable fluorescent signal. The cycle number at which the fluorescent signal is detected is called the threshold cycle, or Ct. The Ct value of a reaction is determined mainly by the amount of template present at the start of the PCR reaction. Thus, the difference in the template amount between two samples at the starting point can be determined by the difference in Ct value, which is ΔCt.

To analyze the mRNA changes, qPCR was performed using cDNAs of the control (100% germination), NA (0.5% germination), and AA (0.5% germination) seeds. The ΔCt values (Ct of aged sample – Ct of unaged sample) of 29 genes (Table 2.2) for AA and NA seeds are shown in Table 3.5. When the ΔCt values were plotted against the cDNA coding region length (Figure 3.7), there was a good correlation between the ΔCt value and fragment length for both AA (Figure 3.7A) and NA (Figure 3.7B) seeds ($R^2 = 0.6507$ and $P < 0.0001$ for AA data, and $R^2 = 0.7167$ and $P < 0.0001$ for NA data). That is, for a longer fragment, there is a greater difference in Ct values between the aged and unaged seeds, indicating that there was a lower amount of template in the aged seed sample likely due to degradation during aging. Remarkably, the slopes of the regression equations for AA and NA seeds in Figure 3.7A and 3.7B were very similar (0.0020 and 0.0019, respectively), indicating a similar linear relationship between the ΔCt value and length of stored mRNAs in NA and AA seeds with the same percentage of germination. Also, for one gene (B4), the qPCR data showed a decrease in both NA and AA seeds (Table 3.5), which the RT-PCR analysis could not clearly detect. These qPCR results further showed that the levels of all stored mRNAs analyzed decreased in aged seeds and the extent of decrease is correlated with mRNA length.

Table 3.5 △Ct values for 29 genes in qPCR analysis on changes in stored mRNAs in aged Arabidopsis seeds.

Gene code	Gene ID	Size (bps)	Average Ct value ± standard error			△Ct	
			Control	AA	NA	AA	NA
A18	At1g74370	786	20.9±0.2	22.1±0.0	21.9±0.1	1.1±0.1	0.9±0.1
A20	At3g01650	1470	22.1±0.1	23.6±0.1	24.2±0.2	1.6±0.1	2.2±0.2
A21	At5g14420	1407	24.1±0.2	25.6±0.1	24.8±0.2	1.5±0.1	0.7±0.2
A26	At5g38895	981	22.0±0.1	23.2±0.6	23.8±0.3	1.2±0.4	1.8±0.1
A27	At5g41350	639	21.4±0.3	22.6±0.1	22.1±0.2	1.2±0.1	0.7±0.2
A31	At1g55530	1056	23.9±0.1	25.6±0.3	24.8±0.2	1.7±0.3	0.8±0.2
A33	At2g17730	762	27.0±0.0	28.4±0.1	27.8±0.1	1.4±0.1	0.8±0.1
A35	At5g47570	1400	26.8±0.1	28.5±0.1	27.9±0.1	1.6±0.1	1.0±0.1
A38	At4g00335	537	25.8±0.2	26.3±0.1	25.8±0.2	0.6±0.1	0.1±0.2
A40	At1g63170	1146	27.3±0.1	28.5±0.1	28.0±0.1	1.2±0.1	0.7±0.1
A41	At1g12760	1227	24.2±0.1	26.8±0.2	25.6±0.2	2.6±0.2	1.4±0.2
A42	At3g61180	1140	26.0±0.0	27.8±0.2	26.9±0.2	1.8±0.2	0.9±0.2
A44	At3G46510	1983	26.2±0.0	27.4±0.3	28.1±0.1	1.3±0.3	1.9±0.1
A45	At1g29340	2190	25.1±0.2	27.3±0.1	27.5±0.1	2.2±0.1	2.4±0.1
A47	At5G67530	1788	28.2±0.3	31.5±0.2	29.7±0.1	3.3±0.2	1.5±0.1
B1	At4g10250	590	24.5±0.3	25.4±0.1	25.4±0.1	0.9±0.1	0.9±0.1
B2	At5g59720	490	27.9±0.5	28.5±0.5	28.4±0.1	0.5±0.5	0.5±0.1
B4	At5g12030	461	15.6±0.0	16.3±0.0	15.9±0.1	0.7±0.0	0.4±0.1
B5	At4g27670	685	28.3±0.3	28.8±0.0	27.3±0.1	0.5±0.0	0.9±0.1
B7	At1g16030	1940	18.8±0.2	23.5±0.2	22.7±0.2	4.7±0.2	3.9±0.2
B9	At2g32120	1692	22.1±0.1	25.6±0.1	24.5±0.2	3.5±0.1	2.3±0.2
B10	At5g09590	2049	24.9±0.2	29.2±0.1	28.0±0.1	4.2±0.0	3.1±0.1
B12	At5g56000	2100	26.8±0.2	30.6±0.5	29.7±0.2	3.8±0.5	2.9±0.2
B13	At2g04030	2343	26.6±0.3	31.4±0.2	30.4±0.3	4.8±0.2	3.8±0.3
B14	At3g07770	2400	28.8±0.2	31.5±0.1	31.9±0.4	2.6±0.1	3.1±0.4
B16	At1g74310	2736	23.2±0.3	29.9±0.1	25.8±0.2	6.7±0.2	2.5±0.1
B17	At5g15450	2907	29.5±0.1	32.8±0.2	32.2±0.1	3.3±0.2	2.7±0.1
B20	At5g51070	2838	25.3±0.3	32.2±0.2	31.0±0.1	6.9±0.2	5.8±0.4
D5	At1g15690	2313	23.7±0.2	28.4±0.3	27.4±0.2	4.7±0.3	3.7±0.2

cDNAs of the unaged (100% germination), acceleratedly aged (AA, 0.5% germination) and naturally aged (NA, 0.5% germination) Arabidopsis seeds were used. A total of 29 genes were used. Each ΔCt value was the average of three biological repeats (with the standard error shown). For each gene analyzed, the ΔCt = Ct of the aged sample – Ct of the unaged sample.

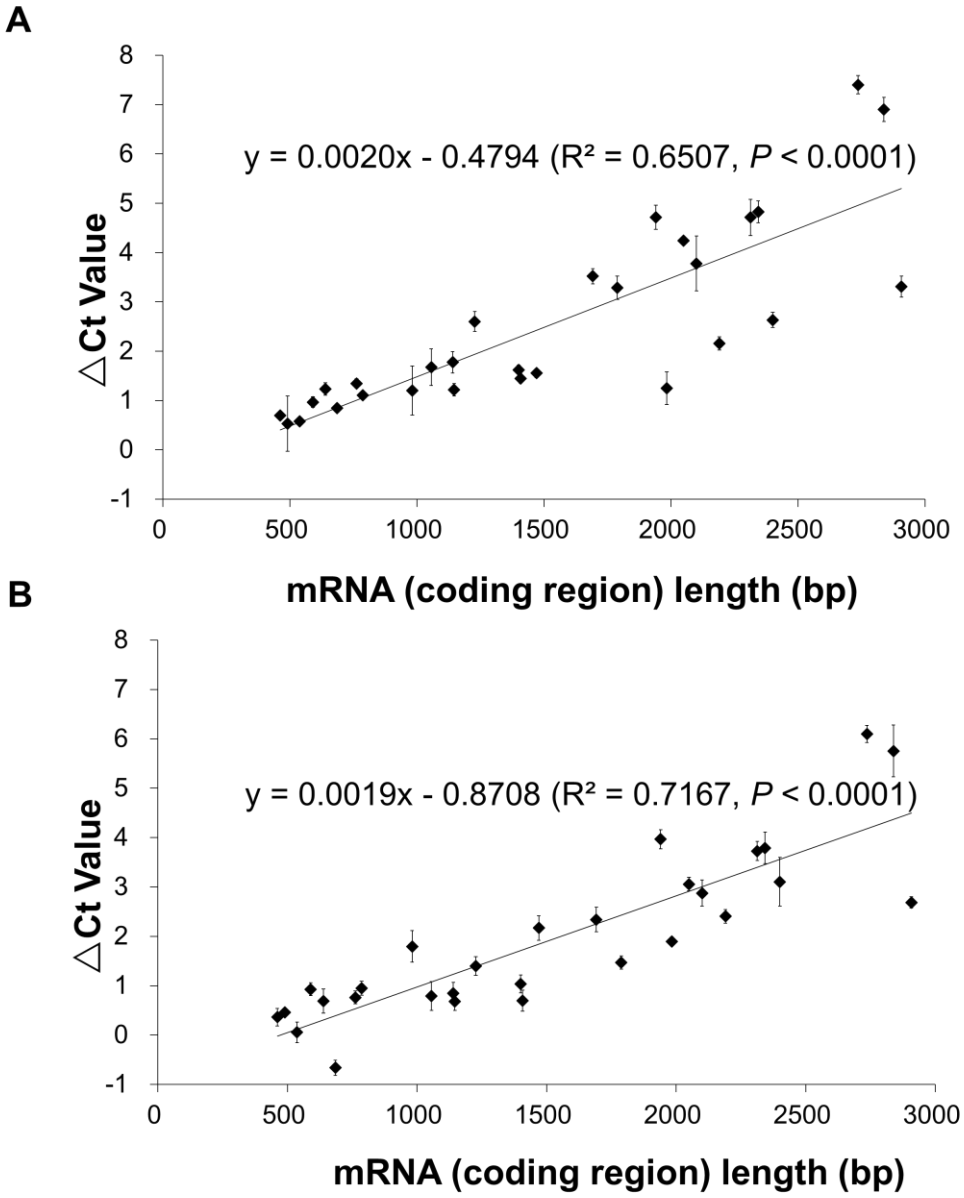


Figure 3.7 The relationship between the length of Arabidopsis stored mRNAs and ΔCt value in qPCR analysis. cDNAs of the unaged control, AA (acceleratedly aged, 0.5% germination) and NA (naturally aged, 0.5% germination) seeds were used. qPCR analysis was conducted for 29 genes which differ in the cDNA (coding region) length. For each gene, the difference in the Ct value between the aged and control samples was obtained as ΔCt = Ct of the aged sample – Ct of the unaged sample. Each ΔCt value was the average of three biological repeats (with the standard error shown). (A) ΔCt values of 29 genes between the AA and unaged control seeds, and the correlation with the cDNA length. (B) ΔCt values of 29 genes between the NA and unaged control seeds, and the correlation with the cDNA length. The linear regressions were produced with Excel 2010 and both have a $P < 0.0001$.

3.2.6 Analysis of the relationship between ΔCt value and mRNA fragment length in aged seeds

Despite a good correlation between ΔCt value and cDNA fragment length, deviations still exist for many genes. The genes and primers (Table 2.2) used in the initial analysis were from our previous studies for different purposes, and might vary in the annealing temperature and PCR amplification efficiency. Further, different mRNAs might differ in the speed of degradation. Thus, to better determine the relationship between mRNA fragment length and the decrease in the mRNA level, we used different fragment lengths of the same stored mRNA and primers with annealing temperatures about 60 °C for better specificity and consistency in amplification. Stored mRNAs of two genes, B16 and B20, with coding regions longer than 2 kb were analyzed using the control, NA (0.5% germination), and AA (0.5% germination) seeds. cDNA regions of 250, 500, 1000, 1500, 2000 and 2500 bp (starting from the STOP codon and counting towards the 5'-direction of an mRNA, Figure 3.8 and Table 2.2) were analyzed, with three different RNA samples for either NA or AA seeds. The specific primers used here showed good specificity in PCR amplification as evident by a single and sharp peak in the qPCR melt-curves.

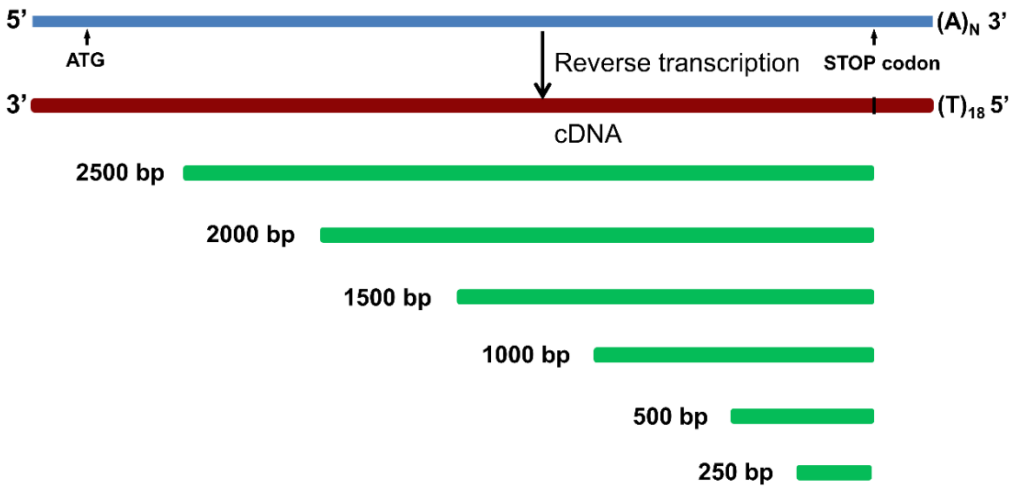


Figure 3.8 Diagram to show the positions of different fragments on the cDNA derived from a stored mRNA. First-strand cDNAs were synthesized from stored mRNAs with an oligo (dT)18 primer. Fragments of different lengths start from the STOP codon and extend towards the 5'-direction of a stored mRNA.

For both B16 and B20, there was a very tight correlation between the ΔCt value and cDNA fragment length (with an R^2 value > 0.90 and $P < 0.0001$; Figure 3.9), clearly indicating that the chance for a damage or lesion to occur to an mRNA fragment increases with the fragment length. The increased ΔCt value for larger fragments could not be due to a difference in cDNA amount or quality since the same cDNA sources were used for amplifying different fragments of the same gene. The difference must be due to the damage or degradation of stored mRNAs. The longer the fragment, the higher the chance for the damage to occur at any point along the entire length of that fragment. If such a damage occurs to a transcript, no full-length cDNA template will be produced and available for PCR amplification.

It is also interesting to observe that, although the slopes were similar for AA and NA seeds, the slope for NA was slightly flatter than that for the AA seeds. Since both seeds had a germination rate of about 0.5%, there was slightly more degradation of stored mRNAs for the two genes in the AA seeds. This ΔCt difference between NA and AA seeds is reminiscent of a difference in the DNA band brightness observed in RT-PCR analysis (Figure 3.6).

3.2.7 Analysis of the relationship between the decrease in stored mRNAs and seed germination

Since traditionally seed aging status is assessed by germination, the relationship between the decrease in stored mRNAs (in ΔCt value) and seed germination was determined. The above results showed that the decrease in stored mRNAs in aged seeds could be clearly analyzed by qPCR, and longer fragments are more sensitive to aging. First, NA seeds were used. These seeds have been collected in our lab at various times and stored at 4 ^{o}C. Since they were collected at different times, variation in the parental plants and seed quality could affect the speed of seed aging. Their aging status was thus based on seed germination percentage rather than storage duration. For the qPCR analysis, the 2000 bp fragment of B16 was used since a longer fragment showed a greater change in ΔCt in aged seeds and thus is more sensitive to seed aging. The qPCR results indicated that the ΔCt value increased with decreased seed viability, but followed a sigmoidal pattern (Figure 3.10A). Similar results

were found with AA seeds (Figure 3.10B).

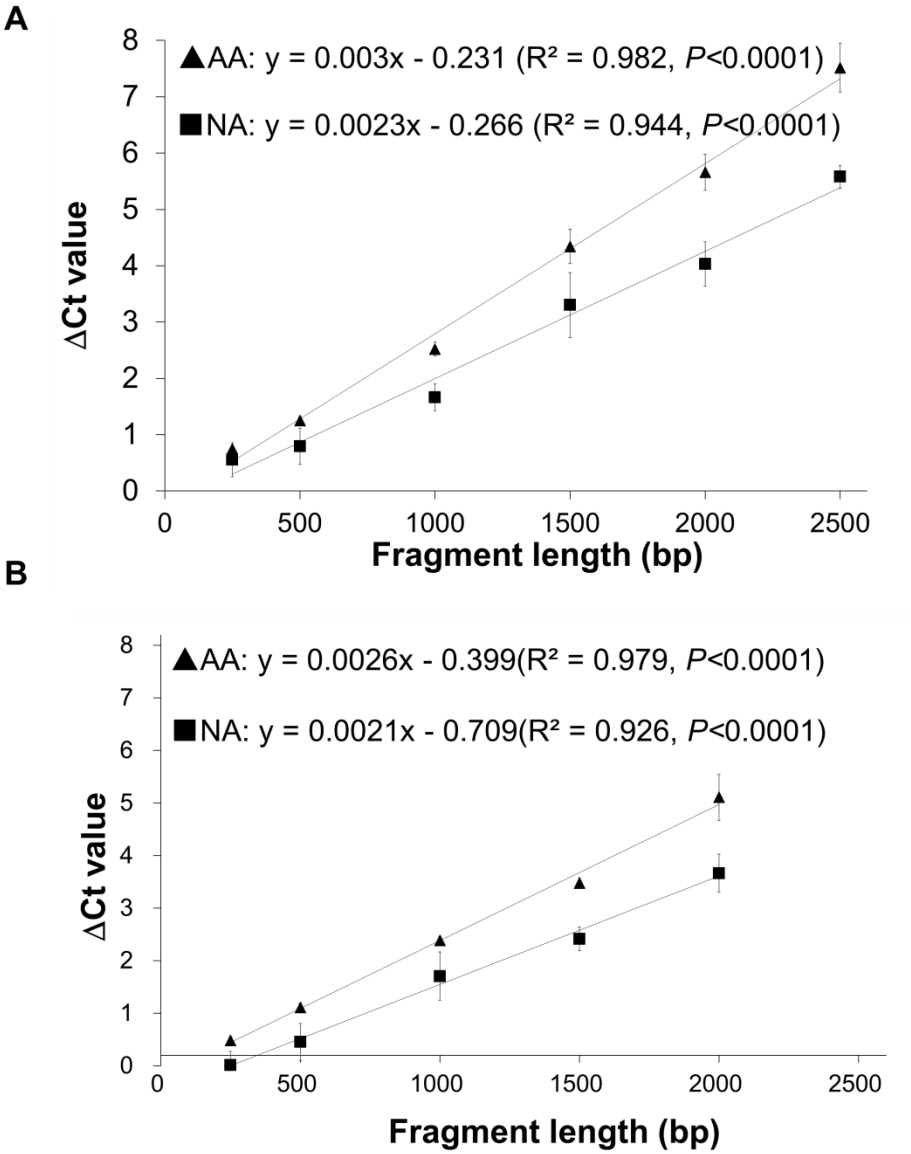


Figure 3.9 The relationship between ΔCt value and cDNA fragment size analyzed using the same genes. cDNAs of the unaged (100% germination), AA (acceleratedly aged, 0.5% germination) and NA (naturally aged, 0.5% germination) seeds were used to analyze different fragments of the B16 and B20 genes. Each ΔCt value was based on three biological repeats with the standard error shown. (A) Correlation between the ΔCt value and cDNA fragment length for six cDNA fragments (starting from the STOP codon) of the gene B16 (with the six fragments being 250, 500, 1000, 1500, 2000 and 2500 bp). (B) Correlation between the ΔCt value and cDNA fragment length for five cDNA fragments (starting from the STOP codon) of the gene B20 (with the five fragments being 250, 500, 1000, 1500, 2000 and 2500 bp). Linear regressions were obtained with Excel 2010 and all have a $P < 0.0001$. ▲: ΔCt values for AA seeds, ■: ΔCt values for NA seeds.

Successful germination is the result of complex processes involving many biomolecules (Almoguera *et al*., 2009; Bueso *et al*., 2014; Chen *et al*., 2012; Rajjou *et al*., 2004). It represents a watershed event (meaning whether a seed can or cannot germinate). On the other hand, changes or damage may be constantly occurring to macromolecules such as stored mRNAs during the seed aging process. Thus, the sigmoidal relationship between seed viability and the ΔCt value is likely because the seed viability kinetics is non-linear during seed aging process.

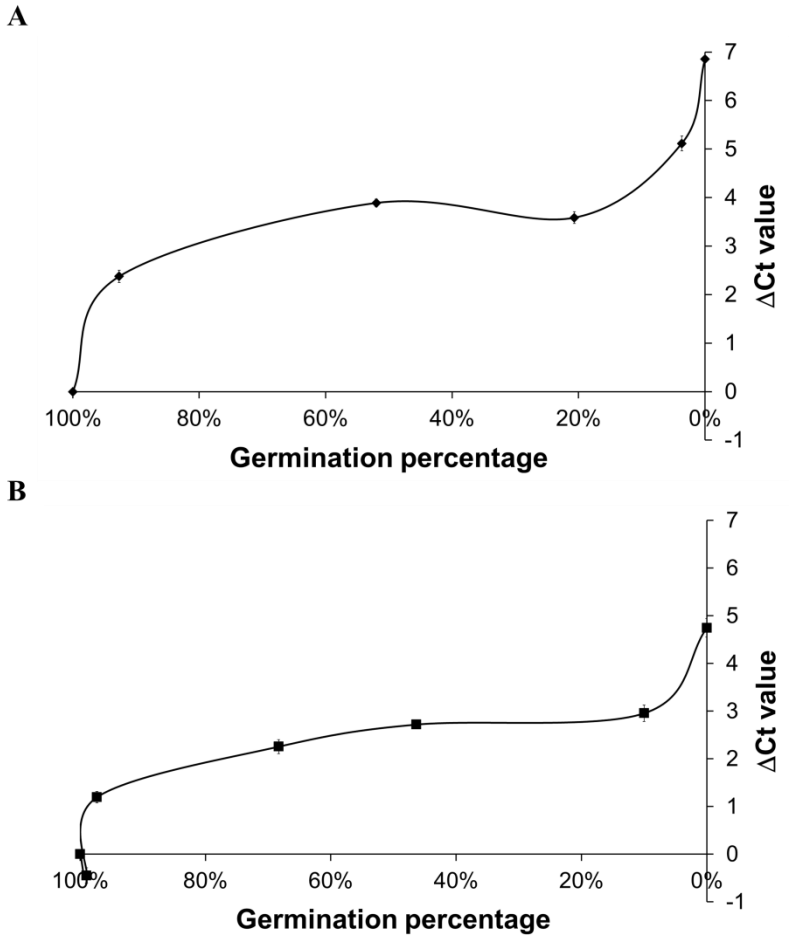


Figure 3.10 Relationship between ΔCt for the cDNA (2500 bp fragment) of gene B16 and Arabidopsis seed germination percentage. cDNAs from naturally aged (NA) and acceleratedly aged (AA) seed samples with different percentages of germination were used. Each datum is based on three technical replicates with the standard error indicated. (A) Relationship between the ΔCt value and germination percentage of six AA samples (with the germination percentage being: 100%, 92.7%, 52%, 20.7%, 3.7% and 0%). (B) Relationship between the ΔCt value and germination percentage of five NA samples (with the germination percentage being: 100%, 99%, 97.3%, 68.3%, 46.3% 10% and 0%).

3.2.8 Development of a quantitative method for determining the changes in stored mRNA levels during seed aging

To better determine the mRNA changes with seed aging time, seeds with similar quality and constant aging conditions were necessary. Thus, a set of seeds from the same seed source and acceleratedly aged for different lengths of time were used. The seed viability of the AA seeds are shown in Figure 3.11. It is clear that the seed viability decreased non-linearly. Under the aging conditions used, for the first six days the change was very slow and for the next six days the germination percentage decreased from above 80% to about 10%. After 14 days, the seeds completely lost the ability to germinate. This seed germination curve fits well with the generalized model of seed aging, consisting of (1) an asymptomatic phase during which the ability of seeds to germinate changes relatively little, (2) a phase of "rapid mortality", and (3) the last phase in which seeds could no longer germinate (Walters *et al*., 2010).

Since the aging status (seed aging time and seed viability) of the AA seeds was clearly known (Figure 3.11), the AA seeds were used to further evaluate the relationship between ΔCt value and aging time. Also, since the ΔCt value from qPCR analysis depended on the

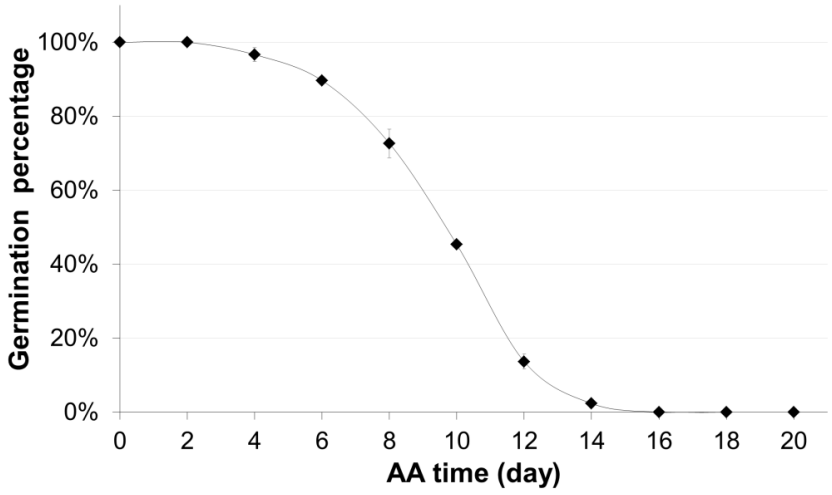


Figure 3.11 Change in the percentage of germination for Arabidopsis seeds following the accelerated aging treatment. Unaged seeds were treated with AA conditions for the indicated times. Each datum is based on three replicate plates with 100 seeds each and the standard errors are shown..

fragment length, we compared different mRNA fragments (starting from the STOP codon position) of B16 and B20. The ΔCt values for different fragments correlated with the aging time and interestingly the correlation coefficient increased with the cDNA fragment length (Figure 3.12). For the 250 bp fragments of both genes, the lines were relatively flat and correlations with the aging time were lower. The correlation coefficients for fragments longer than 1500 bp were close to or greater than 0.90. The slope of the ΔCt regression lines also increased gradually with increasing fragment length. However, the increases were not exactly even as the fragment length increased from 500 to 1000, 1500, 2000 and 2500 bp. It is possible that some regions of mRNAs may be more prone to damage. Another possible factor is that for different fragments, different pairs of primers were used, which could have different efficiencies in PCR amplification. Nevertheless, the ΔCt values for all long fragments (1500 bp or longer) correlated highly with the aging time.

Despite the high correlation coefficients, there were still some deviations from the linear lines in Figure 3.12A and B. Apart from possible experimental variation, another factor that could cause sample to sample variation is differences in cDNA template concentration. Even though the same amount of RNA was used in synthesizing cDNA, any difference in spectrophotometric measurement and any step during cDNA synthesis could cause a difference in cDNA concentration among different samples. To correct this difference, an internal control would be needed. In qPCR analysis, to minimize the differences in template amounts, a gene with a constant expression level is used as an internal control to normalize the Ct value (Livak and Schmittgen, 2001). Since all stored mRNAs analyzed undergo degradation, an mRNA different from the one being analyzed would not serve as a good reference for normalization, and thus a different approach is needed. Because the level of a shorter fragment of mRNA changes much less than a longer fragment, it can be reasoned that a relatively short fragment could be used as an internal reference to normalize the Ct value of a larger fragment. In our analysis, a 250 bp fragment was used for normalization. This approach is similar to the ΔΔCt method (Livak and Schmittgen, 2001) except that a short fragment from the same mRNA is used as the reference.

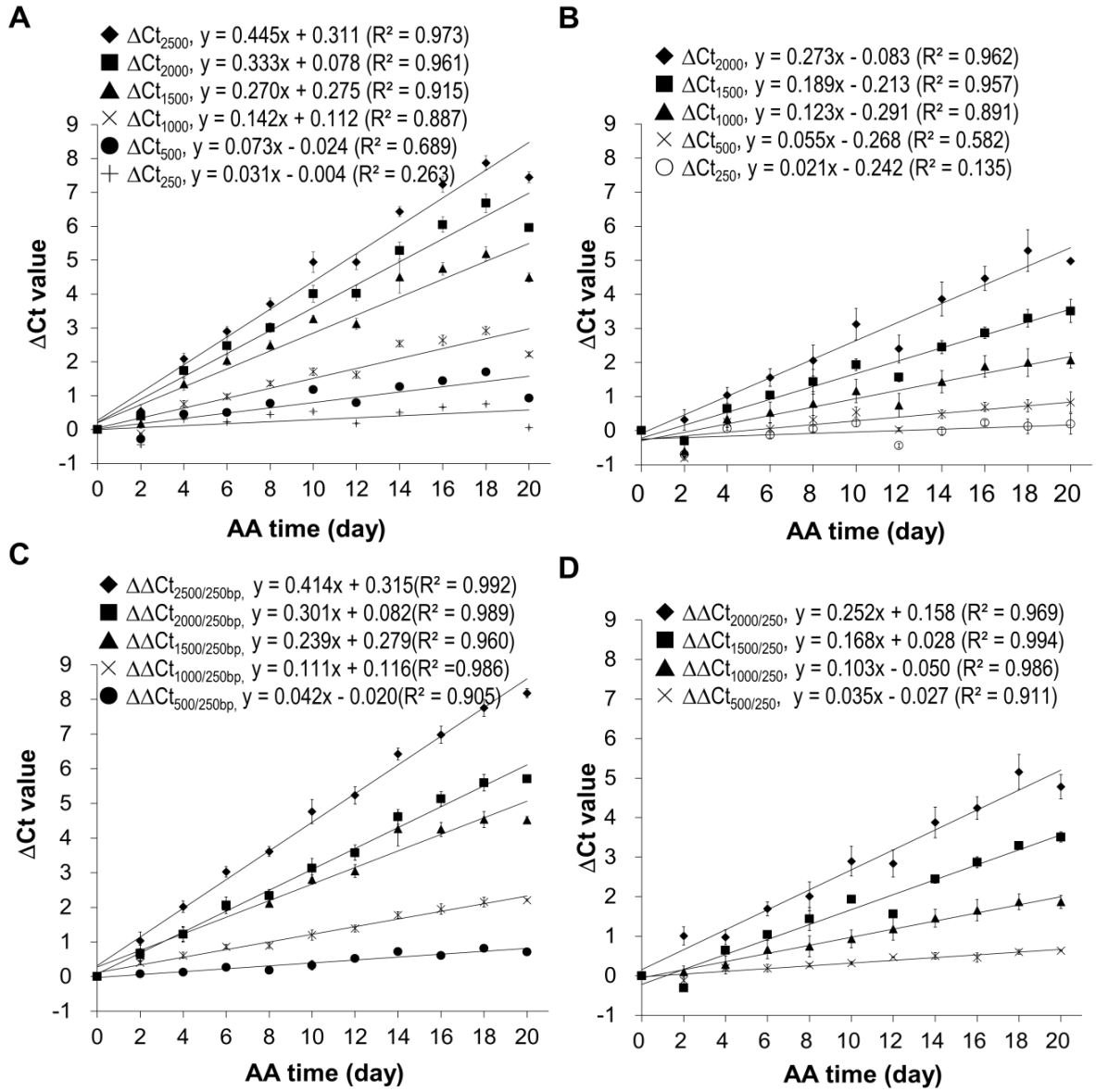


Figure 3.12 Correlation of mRNA degradation with seed aging time analyzed using different cDNA fragment lengths of two genes. cDNAs of the unaged control and ten acceleratedly aged (AA) seed samples for the indicated aging time (day) were used. Different fragments of the gene B16 (250, 500, 1000, 1500, 2000, and 2500 bp) or B20 (250, 500, 1000, 1500, and 2000 bp) were used. ΔCt values are based on three technical repeats. (A) Correlation between ΔCt values for six different fragments of B16 and aging time. The subscript after ΔCt indicates the cDNA fragment length (as ΔCt_{2500} indicates the ΔCt is for the 2500 bp fragment). (B) Correlation between ΔCt values for five different fragments of B20 and aging time. (C) The correlation between normalized ΔCt values for six different fragments of B16 and aging time. The normalized ΔCt or $\Delta\Delta Ct_{2000/250bp} = \Delta Ct_{2000bp} - \Delta Ct_{250bp}$. (D) The correlation between normalized ΔCt values of different fragments of B20 and aging time. Linear regression functions were generated by Excel 2010 and all have a $P < 0.01$ (except for the 250 bp fragment of B16 and B20).

Using the B16 2000 bp fragment as an example, the ΔCt value was calculated from the Ct value of the aged sample minus the Ct value of unaged control resulting in $\Delta Ct_{2000(a-c)}$. Similarly, the ΔCt value for the short 250 bp fragment was calculated from the perspective Ct values for aged and unaged samples, resulting in $\Delta Ct_{250(a-c)}$. The normalized ΔCt (or ΔΔCt) was obtained as follows: $\Delta\Delta Ct_{2000/250} = \Delta Ct_{2000(a-c)} - \Delta Ct_{250(a-c)}$. In these ΔCt designations, the fragment lengths are indicated since the ΔCt value depends on the fragment length, and subscripts 2000/250 indicate the 2000 bp fragment being normalized with the 250 bp fragment. The ΔΔCt values were calculated using the data presented in Figure 3.12A and 3.12B. As shown in Figure 3.12C and 3.12D, the correlations improved for all fragments, showing that normalizing the ΔCt values using a much shorter fragment reduced data variation.

It should be pointed out that the long and short fragments do not need to be fixed as 2000 bp and 250 bp. Other lengths can be used similarly. From Figure 3.12, it is clear that the slopes of the ΔCt curves depend on fragment length. Thus, if the slope of the long fragment is linear and the rate of change is much more than that of a short fragment, the long fragment would be better and more sensitive for analyzing the changes in the ΔCt value. For the length of the long fragment, one limitation is the Taq polymerase used in qPCR which is not very good for amplifying very long fragments.

To assess changes in different stored mRNAs, mRNAs of six genes were analyzed using a 2000 bp fragment. As shown in Figure 3.13, the unnormalized ΔCt value correlated well with aging time for all six genes (Figure 3.13A). Still, after the normalization, the correlation coefficients were further improved (Figure 3.13B). These results show that the normalized ΔCt values have improved the correlation with the seed aging time. The minor differences of slope can be explained by possible differences in the primer binding or mRNA degradation rate. Due to the very good correlations, any of the stored mRNAs analyzed here could serve as a biomarker to assess seed aging.

3.2.9 Estimating the relative amount of undamaged stored mRNAs during seed aging

Although the ΔCt (or ΔΔCt) values highly correlate with seed aging time, they do not

show directly the changes in the level of an mRNA during seed aging. We derived the following equation to estimate the relative amount of undamaged stored mRNAs based on the qPCR data (see equation derivation in Methods):

$$N_a / N_c = ½^{\Delta Ct(a-c)} \quad (9)$$

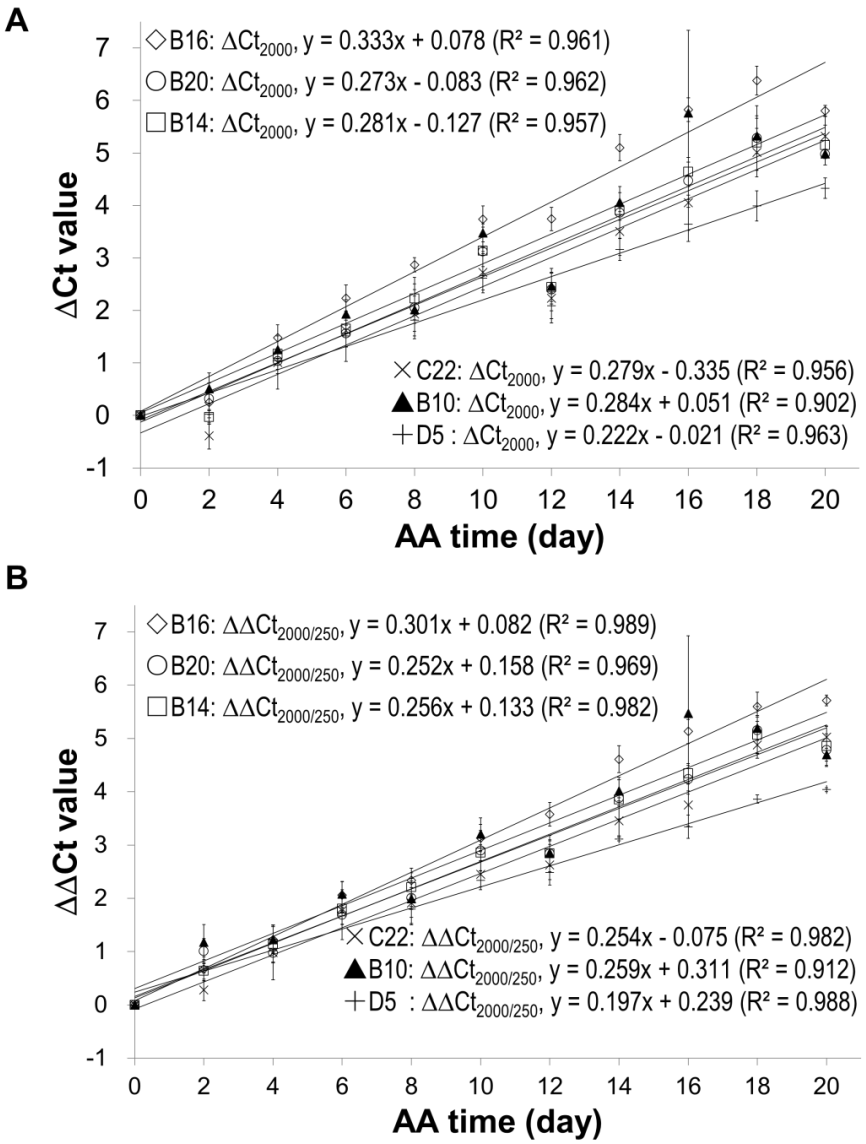


Figure 3.13 The relationship between the ΔCt values from analyzing stored mRNAs of six genes and seed aging time. cDNAs of the unaged control and ten acceleratedly aged seeds for the indicated time (days) were used. For each of the six genes (B16, B20, B14, C22, B10, and D5), a 2000 bp cDNA fragment was analyzed. ΔCt = Ct of the aged sample – Ct of the unaged control. All ΔCt values are the averages of three biological repeats with standard errors shown. (A) Relationship between the ΔCt value of the 2000 bp fragment (ΔCt_{2000}) and AA aging time for the six genes. (B) Relationship between the normalized ΔCt value of the 2000 bp fragment ($\Delta\Delta Ct_{2000/250}$) and AA aging time. $\Delta\Delta Ct_{2000/250bp} = \Delta Ct_{2000bp} - \Delta Ct_{250bp}$. Linear regressions were obtained with Excel 2010 and all have a $P < 0.0001$.

Where

N_a = cDNA copy number of the aged sample

N_c = cDNA copy number of the unaged control

For instance, for $B16_{2000/250}$, the percentage of undamaged mRNA in aged seeds could be estimated by substituting "ΔCt(a-c)" in the above equation with the linear regression function for $B16_{2000/250}$, as follows:

$N_a / N_c * 100\% = ½^{0.301x + 0.082} * 100\%$

where *x* refers to AA time in days

Thus, the percentages of undamaged $B16_{2000/250}$ and $B20_{2000/250}$ templates in aged seed samples were estimated based on the ΔCt values from Figure 3.13B and showed an exponential decrease over the aging time (Figure 3.14).

3.2.10 Half-life time for different fragments

We have also developed a formula for estimating the half-life ($T_{1/2}$) of the stored mRNA, when $N_a = ½ N_u$:

$N_a / N_u = ½ = ½^{\Delta Ct}$

The half-life time is when ΔCt = 1. For instance, the half-life time ($T_{1/2}$) for $B16_{2000/250bp}$ fragment can be calculated by substituting ΔCt with the linear regression of ΔCt values for $B16_{2000/250bp}$:

$\Delta Ct = 1 = 0.310x + 0.215$

$x = T_{1/2} = 2.54$ (days)

Here, since ΔCt values of $B16_{2000bp}$ were normalized by the ΔCt values of $B16_{250bp}$, the $T_{1/2}$ value actually reflects the half-life of $B16_{1750bp}$ (2000 bp – 250 bp). Similarly, the half-life times for all mRNA (or cDNA) fragments in Figure 3.12 C and D were calculated and shown in Figure 3.15. It needs to be pointed out that $T_{1/2}$ here is the time when half of the mRNAs are still "intact" and can produce cDNAs detectable by PCR while the other half could not produce detectable cDNAs presumably due to degradation of the mRNA molecules occurring at least once within the region being amplified by PCR.

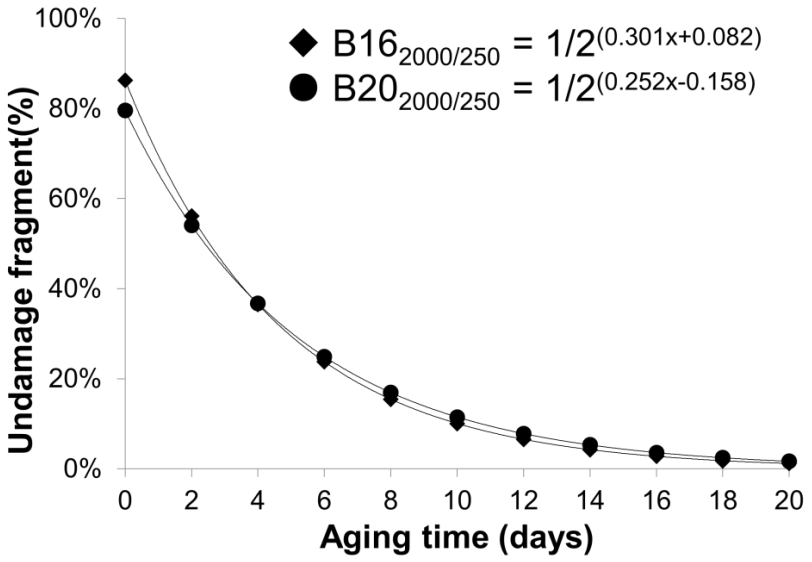


Figure 3.14 The estimated percentage of undamaged mRNAs of B16 and B20 in seeds with different accelerated aging time (days). The mathematical equation (9) was derived to estimate the relative amount or percentage of undamaged stored mRNAs based on the ΔCt regression functions. The regression functions for different fragments of B16 and B20 are from Figure 3.12C and 3.12D. For $B16_{2000/250bp}$, the percentages at different aging times were calculated based on the equation: $y = ½^{\Delta Ct(B16_{2000/250bp})} \times 100\% = ½^{(0.301+\ 0.082)}100\%$, while for $B20_{2000/250bp}$, the equation is $y = ½^{\Delta Ct(B20_{2000/250bp})} \times 100\% = ½^{(0.252+\ 0.158)}\ 100\%$.

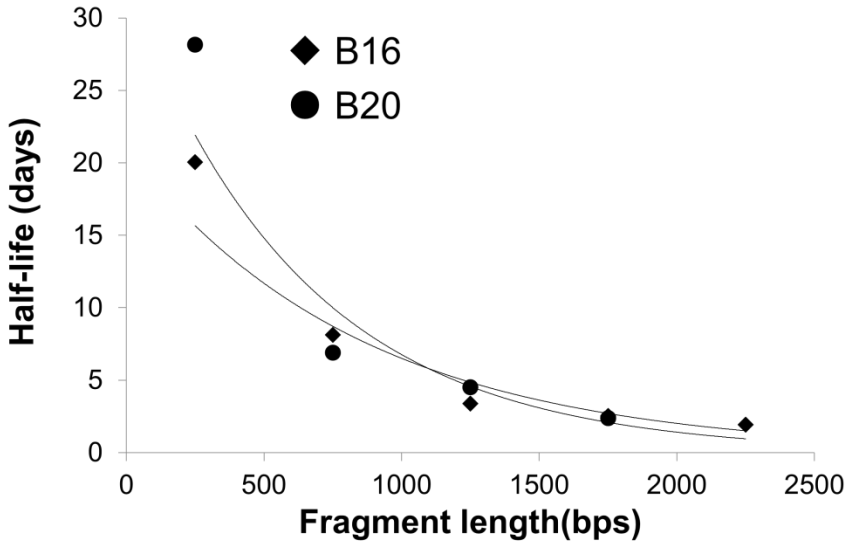


Figure 3.15 The estimated half-life times of different fragments from $B16_{2250bp}$ and $B20_{1750bp}$. The half-lives were calculated (thus have no error bar) based on the regression functions in Figure 3.13B [e.g. the half-life ($T_{1/2}$) for $B16_{1750bp}$ fragment could be calculated: $\Delta Ct = 1 = 0.310x + 0.213$. $x = T_{1/2} = 2.54$ (days)].

3.2.11 Estimating the rate of stored mRNA degradation at the one nucleotide level

To further understand the characteristics of stored mRNA degradation, a method for estimating the rate of mRNA degradation at the one nucleotide level was derived. As described in Materials and Methods the probability of a mRNA (= cDNA) of "*n*" nucleotides with "0" break after "*t*" days of AA treatment can be estimated with the following equation:

$P(0)= e^{-t\beta n} \approx N_a/N_c = ½^{\Delta Ct}$, seed equation (7) in Materials and Methods

where β is the frequency of break per nucleotide per day. Taking a logarithmic transformation of equation (7), we can estimate β value, as:

$\beta = \ln(2^{\Delta Ct})/tnI$, seed equation (8) in Materials and Methods

The $B16_{2000/250}$ fragment of the 18-day AA treatment can be used as an example. Since the ΔCt value for $B16_{2000/250}$ was normalized with the ΔCt value of $B16_{250}$, it reflects the changes for the fragment length of 2000 bp – 250 bp = 1750 bp, and thus the number of nucleotides n = 1750. The aging time t = 18 days, and the normalized ΔCt value was 5.60. Based on equation (8), we have:

$e^{-\beta \cdot 18 \cdot 1750} = ½^{5.60}$, and

$\beta = 1.23 \times 10^{-4}$

The above β value indicates that the probability for one lesion to occur per nucleotide per day is 1.23×10^{-4} for the B16 fragment under the current AA conditions. Similarly, the β values for other fragments of B16 and B20 and aging times were calculated (Table 3.6). The β values were remarkably similar at different aging times for different fragment sizes of B16 or B20, indicating that the two stored mRNAs degraded at a constant rate over the aging time analyzed. Regarding different stored mRNAs, the mRNAs of B16 and B20 also degraded at similar rates, as shown by the total averages of the β values (1.22×10^{-4} vs 0.94×10^{-4}) (Table 3.6). Further, the β values for the stored mRNAs of six genes (Figure 3.13B) confirm, after an ANOVA test, that they generally degraded consistently over the aging time, and with similar rates per nucleotide compared to each other (Table 3.7).

Table 3.6 Estimated β values for different fragments at different aging times.

Gene code	Fragment[(1)]	Fragment size (bp)	β value (break per nucleotide per day x 10^{-4})[(2)]						
			Day 4	Day 8	Day 12	Day 16	Day 20	Average	Overall
B16	$B16_{2500/250bp}$	2250	1.49±0.22	1.40±0.14	1.35±0.11	1.38±0.09	1.25±0.05	1.37±0.05	1.26±0.00
	$B16_{2000/250bp}$	1750	1.42±0.08	1.27±0.01	1.27±0.01	1.33±0.00	1.17±0.00	1.29±0.03	
	$B16_{1500/250bp}$	1250	1.43±0.11	1.42±0.04	1.36±0.08	1.42±0.02	1.23±0.03	1.37±0.03	
	$B16_{1000/250bp}$	750	1.02±0.29	1.06±0.01	1.10±0.06	1.14±0.05	1.00±0.00	1.07±0.05	
	$B16_{500/250bp}$	250	0.99±0.08	1.14±0.24	1.40±0.09	1.35±0.29	1.20±0.22	1.22±0.09	
	Average		1.26±0.09	1.25±0.06	1.29±0.04	1.32±0.06	1.16±0.05		
B20	$B20_{2000/250bp}$	1750	1.12±0.13	1.09±0.14	0.96±0.12	1.10±0.06	0.96±0.06	1.05±0.05	0.95±0.06
	$B20_{1500/250bp}$	1250	0.85±0.20	1.03±0.21	0.95±0.03	0.95±0.07	0.94±0.01	0.94±0.05	
	$B20_{1000/250bp}$	750	0.60±0.51	0.92±0.27	0.90±0.21	0.95±0.15	0.89±0.10	0.85±0.11	
	$B20_{500/250bp}$	250	1.34±0.82	1.07±0.25	1.08±0.10	0.85±0.27	0.50±0.47	0.97±0.19	
	Average		0.98±0.18	1.03±0.03	0.97±0.04	0.96±0.06	0.82±0.12		

A method (equation 8) was developed to estimate the "probability" of producing a break in the cDNA of a stored mRNA at the one nucleotide level, which is referred to as the β value (see Materials and Methods), using the ΔCt values of qPCR analysis in Figure 3.12C and D. Each β value is based on three technical repeats, and the standard error is indicated.

(1) The fragment analyzed and fragment used for normalizing the ΔCt value are indicated. For instance, $B16_{2500/250bp}$ indicates that a 2500 bp fragment of B16 was analyzed and normalized with the ΔCt of a 250 bp fragment.

(2) β value was estimated using the ΔCt value of the AA seed sample at the indicated aging time, and is the probability of one break per nucleotide per day.

Table 3.7 Estimated β values for stored mRNAs of six genes at different aging times.

Fragment	β value (break per nucleotide per day x 10^{-4})						
	Day 4	Day 8	Day 12	Day 16	Day 20	Average	overall
$B16_{2000/250bp}$	1.42±0.02	1.27±0.01	1.27±0.01	1.33±0.00	1.17±0.00	1.29±0.00	
$B20_{2000/250bp}$	1.12±0.04	1.09±0.04	0.96±0.04	1.10±0.02	0.96±0.02	1.05±0.01	
$B10_{2000/250bp}$	0.83±0.09	0.75±0.05	0.74±0.04	0.83±0.01	1.04±0.02	0.84±0.02	0.98±0.01
$B14_{2000/250bp}$	0.91±0.03	0.89±0.07	0.74±0.03	0.99±0.02	1.01±0.02	0.91±0.01	
$C22_{2000/250bp}$	0.74±0.15	0.77±0.08	0.71±0.05	0.85±0.04	0.99±0.01	0.81±0.03	
$D5_{2000/250bp}$	0.88±0.03	0.78±0.04	0.66±0.02	0.78±0.02	0.87±0.01	0.79±0.01	

The β values for stored mRNAs of six genes at different aging times were estimated based on the ΔCt values of qPCR analysis, as shown in Figure 3.13B. Each β value is based on three biological repeats, and the standard error is indicated.

The β value, independent of the mRNA length, is more useful for the comparisons of stored mRNA degradation and seed aging under different conditions. Since the speed of seed aging depends greatly on the temperature, we determined the β value of seeds aged at different temperatures (22, 30, 33, 37, and 40 ºC) using seeds stored at 4 ºC as the reference. As shown in Figure 3.16, the temperature greatly affected the β value. Interestingly, there seemed to be a turning point around 33 ºC, after which the β value increased much faster. For instance, the β value roughly doubled when the temperature increased by 9 degrees from 22 ºC to 33 ºC, while it more than doubled when the temperature increased by 3 degrees from 37 ºC to 40 ºC (Figure 3.16). These results show quantitatively that stored mRNA degradation accelerates with increasing temperature and provide a mechanistic explanation for the accelerated seed aging at higher temperatures. In addition, the β values of seeds aged for 8 and 16 days at the same temperature were very similar (Table 3.8 and Figure 3.16), further supporting that the frequency of breakdown at the one nucleotide level in a stored mRNA is fairly constant over the aging time.

Table 3.8 Estimated fragment β values for Arabidopsis seeds aged at different aging temperatures.

Gene	AA days	β value (break per nucleotide per day x 10^{-4})$^{(2)}$				
		22 °C	30 °C	33 °C	37 °C	40 °C
$B16_{2000/250bp}$	8	0.27±0.04	0.59±0.09	0.63±0.04	1.29±0.08	2.54±0.21
	16	0.19±0.06	0.46±0.04	0.49±0.05	1.22±0.07	1.86±0.17
Average		0.23±0.05	0.51±0.06	0.57±0.04	1.25±0.07	2.2±0.19

Using the equation (8) in the Materials and Methods, the β values were calculated from the ΔCt values between the control and aged seeds as shown in Figure 3.16. Each β value is based on ΔCt values of three technical repeats, and the standard error is indicated.

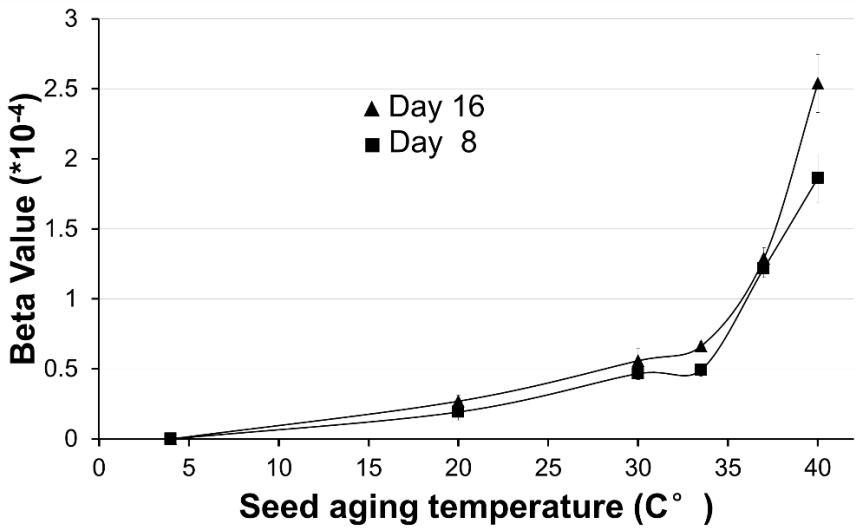


Figure 3.16 Estimated β values for stored mRNAs in seeds aged at different temperatures. Arabidopsis seeds were treated at 22, 30, 33, 37, and 40 °C for 8 or 16 days, with seeds stored at 4 °C used as the unaged control (totally 11 seed samples). Total RNAs and cDNAs were prepared. The 2000 bp fragment of B16 was used in the qPCR analysis, with the 250 bp fragment of B16 used to normalize the ΔCt for the 2000 bp fragment. For each sample, the normalized ΔCt was calculated first, and then the β value was estimated as described in the method. Each β value is based on ΔCt values of three technical replicates with the standard error shown. The seed lot used in this experiment was different from the seed lot used for Figures 3.6 and 3.12.

3.2.12 MinION nanopore sequencing of cDNA derived from seed stored mRNAs

There are over 12,000 stored mRNA species in mature dry Arabidopsis seeds based on genome-wide analyses (Kimura and Nambara, 2010; Nakabayashi *et al*., 2005). Although studies of specific stored mRNAs can reveal how their levels change during seed aging, it is still interesting to understand how seed stored mRNAs change genome-wide and more importantly the differences, if any, among different mRNAs.

A direct way is to perform RNA-Seq using Next Generation Sequencing (NGS) technologies. However, most of the NGS technologies require DNA fragmentation and PCR amplification which may have uneven read coverage (Pelechano *et al*., 2013; Ross *et al*., 2013) and fragment bias (Aird *et al*., 2011). Also the produced short reads require complex data analyses. Pacific Biosciences has a DNA sequence technology based on single molecule DNA synthesis referred to as zero-mode waveguide (ZMW) technology, which produces longer sequence reads than the classical Sanger sequencing technology (van Dijk *et al*.,

2018). However, the cost is higher than some of the short-read technologies such as Illumina technology.

Recently, Oxford Nanopore Technologies introduced a long-read DNA sequencing technology based on nanopore DNA sequencing (Oikonomopoulos *et al*., 2016). A nanopore is a nanoscale pore allowing molecules such as a single strand of nucleic acid to pass through (Maitra *et al*., 2012). The nanopore could be a protein pore inserted into a lipid membrane or a solid nanopore in a solid-state membrane, although the latter technology is still in the developmental stage. When a voltage is applied across the membrane, an ionic current is created, and the current will fluctuate when a strand of DNA passes through. The changes in current, if strong and unique enough, can be utilized for determining DNA sequencing. Thus, nanopore DNA sequencing technology can potentially produce very long reads. The first commercial nanopore sequencing devices of relative small sizes have recently been released by Oxfore Nanopore Technology, and one of them is named MinION.

We used MinION nanopore sequencing to sequence cDNAs of WT and 16-day AA Arabidopsis seeds. The cDNA libraries were prepared according to the modified protocols as described in the Material and Methods.

From the MinION sequencing, I obtained 867,198 and 791,886 events for unaged and AA cDNA library respectively. Among them, 834,652 and 768,432 were converted successfully to nucleotide reads. Of 53,828 reference Arabidopsis transcripts (TAIR10 https://www.arabidopsis.org/), 21,082 were found in our sequence reads. For the comparison between the control and AA samples, only the transcripts that were detected in both libraries were selected, and accordingly the transcript pool was narrowed to 17,523 transcripts.

To determine mRNA degradation, the coverage of the reference transcripts was compared between unaged and AA samples. Since mRNAs were less degraded in unaged seeds, the length of mRNA (or cDNA) for a gene should be longer compared to the aged sample; and the sequencing reads should also be longer and cover a larger sequence span. Since a large number of reference transcripts were only partially covered by the sequencing reads, it is better to use reference transcripts that were well covered by the sequencing reads (i.e. the sequenced reads cover a larger and continuous portion of the reference template) for

comparing the control and AA samples. Thus, if the length of a reference transcript is covered by 75% or more by the sequence reads, it is considered a "highly covered" reference transcript. This threshold is the same as used by Fleming *et al*. (2018b). The sequence read results showed that the unaged sample had 24.64% (4318/17523) of reference transcripts to be highly covered, compared to 12.45% (2181/17523) for the AA sample. When only the highly covered 4318 reference transcripts in the unaged sample were considered, there was a good correlation between the length of reference transcripts and the length of read coverage for the unaged sample [$R^2 = 0.9578$, $y = 0.8993x$, $P < 0.0001$ (Figure 3.17A)], but there was no correlation for the AA seed sample [$R^2 = 0.00$, $y = 0.5421$, $P < 0.0001$ (Figure 3.17B)]. When the 2181 reference transcripts (the highly covered transcripts in AA sample) were considered, high correlations were observed between the length of reference genes and the length of read coverage in the AA sample [$R^2 = 0.9495$, $y = 0.8698x$, $P < 0.0001$ (Figure 3.17C)]. However, the correlation was still high for the unaged sample [$R^2 = 0.6123$, $y = 0.8103x$, $P < 0.0001$ (Figure 3.17D)], which reflected a higher degree of integrity for stored mRNAs in unaged sample, considering the severe decrease in correlation as observed in Figure 3.17A and 3.17B. Therefore, the much lower number of the highly covered genes and the weaker correlation between the length of reference genes and length of read coverage suggest a general degradation of stored mRNA in the AA sample. We further analyzed the mRNA degradation gene by gene. For one gene, the average cDNA length for the aged sample should be shorter than the unaged sample. Thus, if the sequencing reads for one gene are aligned against the reference gene, the shape of the sequence alignment is expected to shift towards the 3' end for the aged sample comparing to the unaged sample. Using B20 as an example, when all sequencing reads of B20 were displayed, the 5' ends of the sequence alignment indeed shifted towards the 3' end for the AA sample, indicating that the lengths of sequencing reads for the AA sample appeared to be shorter on average than those for the unaged sample (Figure 3.17 E).

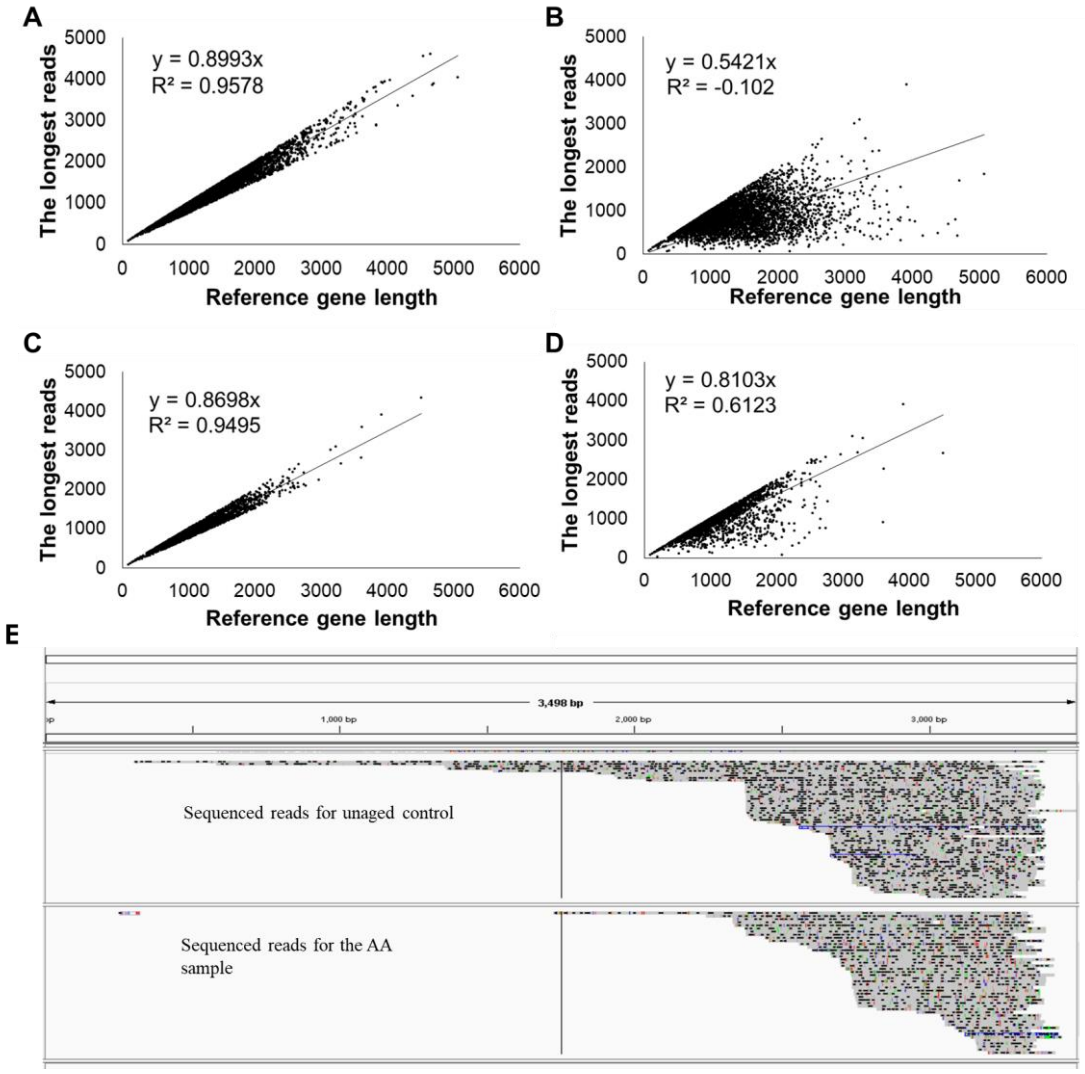


Figure 3.17 Comparison of read coverage between unaged and AA sample. RNA samples from unaged and AA (16-day AA treatment) were used for constructing cDNA libraries, which were sequenced using MinION sequencing. Since mRNAs were less damaged in unaged seeds, the sequencing reads should be longer and cover a higher percentage of Arabidopsis reference transcripts compared to the aged sample. A gene is considered to be "highly covered", if its length is 75% covered by the sequenced reads. (A) The sequenced lengths for the unaged sample were plotted against the reference cDNA lengths using the highly covered 4, 318 reference transcripts in the unaged sample. (B) The sequenced lengths for the aged sample were plotted against the same 4,318 reference transcripts as in (A). (C) The sequenced lengths for the aged sample were plotted against the lengths of the highly covered 2, 181 reference transcripts in the aged sample. (D) The sequenced lengths for the unaged sample were plotted against the lengths of the same 2, 181 reference transcripts as in (C). (E) The sequenced reads for the gene B20 from unaged control and AA sample were aligned to the reference transcript. The alignment was performed with Integrative Genomics Viewer (IGV_Win_2.3.92). The sequenced reads for the unaged control are shown in the upper part, while the sequenced reads for the AA sample in the lower part.

3.2.13 Comparisons with traditional methods of assessing seed aging

In addition to seed germination, other traditional methods have also been used for assessing seed aging. Two common ones are seedling growth (Marcos Filho, 2015) and electrical conductivity (EC) (Matthews and Powell, 2006). The EC test is based on the principle that in water, aged seeds leak more electrolytes than unaged seeds. The cell membrane is a selective permeable membrane, and metabolites inside the cell cannot pass it freely. However, if the membrane is damaged during seed aging, more metabolites will leak out and thus the electrical conductivity of the water solution increases.

For seedling growth, root growth (length) and fresh seedling weight were analyzed. As shown in Figure 3.18, seedling root length (Figure 3.18A) and fresh weight (Figure 3.18B) decreased gradually to zero with seed aging time. However, the decreases in both parameters did not follow an exact linear pattern, but rather a sigmoidal one, similar to that for the germination rate. Further, both parameters could not distinguish seeds aged from 2 to 6 days under the current aging conditions and also failed to detect any difference among the seeds aged for more than 14 days, since no seeds germinated. For the EC test (Figure 3.18C), the data had a correlation with seed aging, but with a much lower coefficient than that of the ΔCt value for the stored mRNAs. Therefore, ΔCt for long fragments of stored mRNAs is a much more accurate method to assess seed aging processes than the classical methods such as seed germination, seedling growth and EC test. Furthermore, two lots of seeds may germinate at 100% but have different vigor, and the ΔCt method for stored mRNAs could determine the difference more accurately.

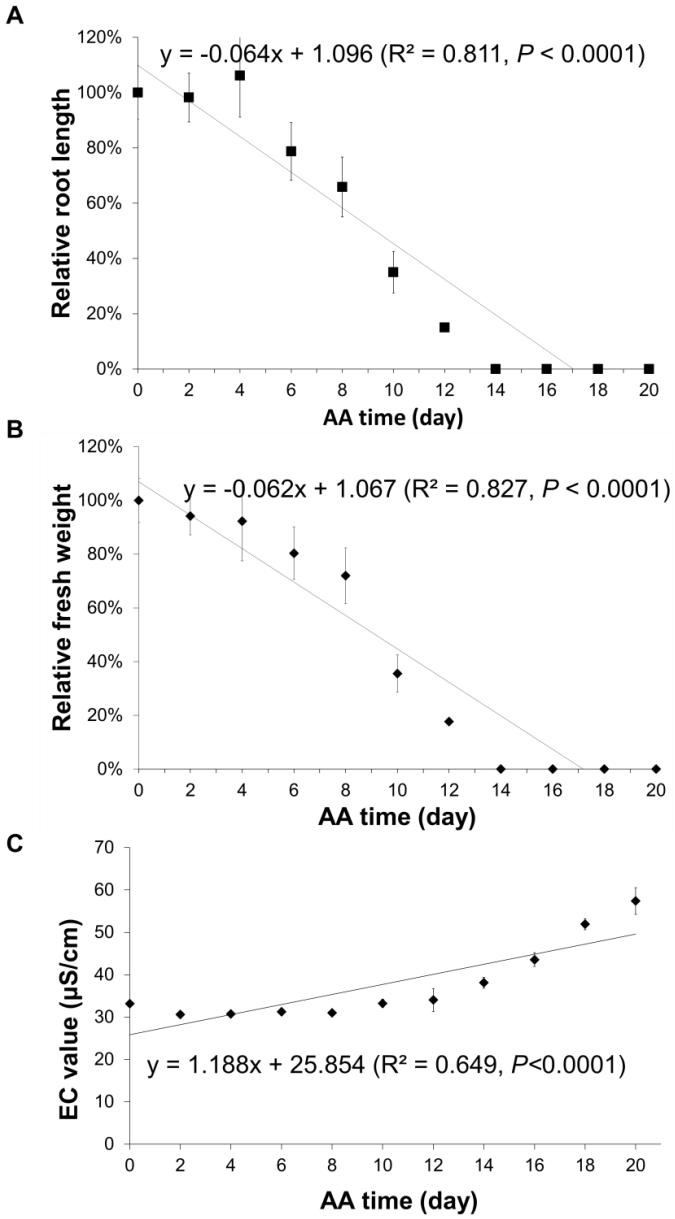


Figure 3.18 Analysis of seed aging by traditional methods. The control and ten acceleratedly aged (AA) seed samples treated for different lengths of aging time (day) were used for seed aging analyses based on seedling growth and electrical conductivity (EC) test. Two parameters of seedling growth were determined. For each time point, five plates were used and each plate had unaged and aged seeds planted side-by-side. After 10 days of incubation, the relative primary root length and fresh seedling weight (the ratio between aged and unaged control samples) were obtained for each plate. Each datum is based on five plates with the standard error shown for figure (A) and (B). (A) Relative primary root length. (B) Relative seedling fresh weight. (C) EC value. Seeds (10 mg) of different aging times (with three biological repeats for each time point) were soaked in 1.5 mL ddH_2O used and EC analysis was performed after 16 hours of incubation. Each datum is based on the three biological repeats with the standard error shown in figure (C). Linear regressions were obtained with Excel 2010 and statistically significant at $P < 0.0001$.

3.3 Discussion

3.3.1 Degradation of stored mRNAs is highly correlated with aging in naturally and artificially aged Arabidopsis seeds

It has been reported that the integrity of rRNAs decreases in aged seeds of several plant species based on the appearance of smaller RNA bands on the electrophoresis graphs (Brocklehurst and Fraser, 1980; Kranner *et al*., 2011; Matthews and Powell, 2006; Reuzeau and Cavalie, 1997; Sharma *et al*., 2018; Thompson *et al*., 1987). However, it was difficult to quantify the relationship between RNA band changes and seed aging time. Recently, Fleming *et al.* (2017) examined the changes in RNA integrity using the RNA integrity number (RIN). The RIN is an index based on multiple features of electrophoresis graphs generated from Agilent Bioanalyzer, including the total RNA ratio, the heights of the 25S and 18S peaks, and the "fast area" (i.e. small RNA molecules/fragments) ratio (to the total area of the electrophoresis graphs) (Imbeaud *et al*., 2005; Schroeder *et al*., 2006). Using soybean seeds stored up to 27 years, a high correlation between the RIN of RNA isolated from cotyledons and the percentage of germination was observed, but the correlation was lower between the RIN of the RNAs from the embryonic axis and percentage of germination (Fleming *et al*., 2017). The authors also extended this analysis to several other plant species and observed similarly that the RNA integrity decreased during seed aging (Fleming *et al*., 2018a).

We also analyzed our Arabidopsis RNA samples using the Agilent Bioanalyzer (Table 3.2). At the start of this study, we had optimized the RNA isolation protocol to ensure the quality and consistency of RNA samples. The RINs were very similar (mostly between 8 – 9) among samples of different aging times (Table 3.2). Thus, for the Arabidopsis RNA samples in this study, the RIN value was not sufficiently sensitive and effective for assessing seed aging compared to the ΔCt value of a stored mRNA. Since the RIN is an integrated parameter based on multiple features of a RNA electrophoresis graphs, it is affected not just by RNA degradation, but also the purity, concentration and abnormal peaks. Furthermore, the method developed here is well suited to analyze the changes of individual RNAs, which

the RIN data could not determine.

Much less research has been done regarding stored mRNAs and seed aging. The transcripts of an Arabidopsis gene, *At3g08030*, and its homologs in two other plant species were found to be present in freshly collected seeds, but reduced or not detectable in aged seeds (Garza-Caligaris *et al*., 2012). The authors suggested that the *At3g08030* transcript could be a potential biomarker for seed aging. However, the quantitative relationship of the *At3g08030* transcript level with aging time or seed germination percentage was not determined. Recently, Fleming *et al*. (2018b) observed fragmentation of stored mRNAs in the soybean embryonic axis by transcriptomic analysis and suggested that mRNA breaks occur at random positions.

The intactness of an mRNA is reflected by its ability to be reverse-transcribed into a cDNA. The cDNAs were synthesized from the polyA tail using an oligo dT primer and we typically use the STOP codon as the start point (Figure 3.8). Since the relative amount of mRNA can be determined by qPCR using cDNA as the template, the difference between the Ct of an aged sample and Ct of the unaged control would reflect the change in the amount of the undamaged mRNA for the aged sample relative to the control.

Our initial RT-PCR analysis showed that almost all stored mRNAs initially analyzed (over 80 of them) showed a gradual decrease with longer stored mRNAs decreasing faster. The results from the subsequent qPCR analysis showed that the degradation of Arabidopsis seed stored mRNAs as represented by the ΔCt value was highly correlated with the aging time. In addition, it was observed that the degradation of stored mRNAs was very similar in AA and NA seeds with the same germination percentage (0.5%) suggesting that the AA seeds aged in a similar way to the NA seeds even although the degradation of stored mRNAs in NA seeds took a much longer time (years) to occur. These results collectively indicate that the degradation of stored mRNAs is highly correlated with seed aging time in naturally and artificially aged Arabidopsis seeds.

3.3.2 A new approach of normalizing qPCR data improves the analysis of stored mRNAs

In qPCR analysis, a reference gene is typically used to normalize the ΔCt value (Livak and Schmittgen, 2001). Since all stored mRNAs showed degradation in seed aging, we used a short fragment (e.g. 250 bp) for normalization. The results show that this approach was effective and improved the correlations between the ΔCt value and aging time for different fragment lengths (Figure 3.12) and for different genes (Figure 3.13).

This method should be applicable for studying stored mRNAs in different plants. The long and short fragments do not have to be the same in lengths as the ones used here. For a different application, it would be good to determine the quantitative relationship of ΔCt value to the fragment length and aging time (Figure 3.12), based on which suitable long and short fragments could be selected.

3.3.3 Quantitative methods are developed to estimate the relative amount of intact mRNAs and rate of mRNA degradation at the one nucleotide level

Results from this study have provided support for two fundamental characteristics of stored mRNA degradation during seed aging. First, the strong linear relationship between the fragment length of a stored mRNA and the ΔCt value when analyzed at the given time point (Figure 3.9) finding indicates that when the mRNA length increases by the same increment (e.g. 500 bp), its level decreases by a similar extent. These results suggest that the damage or degradation of a stored mRNA occurs randomly along the length of a stored mRNA. Second, when a given length (e.g. 2000 bp) of a stored mRNA was considered, the ΔCt value was highly correlated with seed aging time (Figure 3.12 and 3.13), indicating that the time for the mRNA level to decrease by 50% (ΔCt = 1, assuming 100% PCR amplification efficiency) is constant. Based on these two characteristics, we can estimate the relative amount of undamaged mRNAs at a given aging time and frequency of degradation at the one nucleotide level.

The relative amount of undamaged mRNA level could be estimated by substituting ΔCt(a-c) in the equation $N_a / N_c = ½^{\Delta Ct(a-c)}$ with the linear regression function for the stored

mRNA or a given fragment (see Material and Methods). The percentages of estimated undamaged mRNAs provide us with a clearer understanding regarding how stored mRNA levels change over aging time (Figure 3.14).

Meanwhile, the average frequency of "breaks" per nucleotide per day or β value could be calculated using the formula ($\beta = \ln(2^{\Delta Ct})/tn$). A similar approach has been used to estimate the average lesion per DNA strand in DNA damage analysis (Ayala-Torres *et al*., 2000; Yakes and VanHouten, 1997). The similar β values for different fragments of the same stored mRNA (Table 3.6) and for different stored mRNAs (Table 3.7) indicate that a stored mRNA is degraded at a fairly constant rate under the present aging conditions. The β values also suggest that the rates of degradation for six different mRNAs were very similar.

To our knowledge, there has been no previous report on quantifying the degradation of stored mRNAs during seed aging. Understandably, the β value depends on specific aging conditions. Indeed, our analysis of Arabidopsis seeds aged at different temperatures revealed that the β value increased with increasing temperature (Figure 3.16). The β value thus provides a new, simple and quantitative parameter for analyzing stored mRNA degradation and seed aging. It should allow comparisons of different mRNAs, different regions of the same mRNA, and different aging conditions in plant seeds.

Our quantitative analyses of randomly selected stored mRNAs showed that each of the stored mRNAs was degraded at a constant rate over the aging time, while several different stored mRNAs analyzed are also similar in their rates of degradation. Since all stored mRNAs analyzed in this study showed similar trends of gradual decreases, collectively these results suggest a scenario that the majority of stored mRNAs could be degraded with a similar pattern and with a similar rate during seed aging. Given that there are at least 12,000 stored mRNAs in Arabidopsis seeds (Nakabayashi *et al*., 2005), an interesting question is whether differences exist among stored mRNAs in term of the rate of degradation. The scale of work prevented us from comparing the rate of degradation for a large number of individual stored mRNAs in this study. The methods developed in this study provide useful and effective means to address this question through comparative and quantitative analyses of individual stored mRNAs, which had been difficult to determine using existing methods.

Another interesting question is how stored mRNAs are degraded during seed aging. The following considerations lead us to suggest that the breakdown of stored mRNAs is mainly through non-enzyme-based mechanisms. In dry seeds there is very limited free water and thus enzymes and substrates have much less freedom to move around. These conditions together with enzyme denaturing over time determines that the rate of an enzyme-catalyzed reaction would slow down with time. The constant rates and random nature of stored mRNA degradation observed over a long time (under AA and NA conditions) are likely to be due to non-enzyme-based reactions. A similar suggestion has been made by Fleming *et al*. (2018b).

3.3.4 MinION sequencing results also show the degradation of stored mRNAs in seed aging

The MinION sequencing showed that both the number and length of sequence coverage for the "highly covered" reference transcripts were reduced in the AA sample compared to the unaged control, supporting that stored mRNAs are degraded in Arabidopsis seeds during aging. This observation is consistent with what was reported in soybean seeds by Fleming *et al*., (2017).

However, the approach to analyze stored mRNA length changes by MinION sequencing has problems compared to the qPCR approach. First, in constructing the cDNA library, cDNAs needed to be amplified by PCR and then the amplified double-stranded DNA fragments were purified for library construction. Both technical steps could introduce bias and change the relative abundance of different fragment lengths, since the efficiencies of PCR amplification and DNA purification could vary depending on DNA fragment length. Second, shorter cDNA fragments would have a better chance to be fully sequenced than longer cDNA fragments, reducing the representation of long cDNA fragments in a sample. Third, it is more difficult to quantify the average change in fragment lengths based on the sequencing data and correlate fragment length change with aging time. Fourth, a mathematical equation was derived to estimate the frequency of breakdown at one nucleotide, as we did for the qPCR data. Fifth, sequencing entire cDNA libraries requires much more effort, time and cost than analyzing a few mRNAs by qPCR. For these reasons, I did not

pursue sequencing analysis any further.

3.3.5 Stored mRNAs can serve as more precise biomarkers for monitoring seed aging

In assessing seed aging status, the classical methods such as seed germination percentage and seedling growth have a few major problems: little change during the asymptomatic phase and the lack of strictly linear relationship between the parameter analyzed and aging time. A better parameter should have a tight linear relationship with seed aging time and be able to detect changes during the asymptomatic phase (Fu *et al*., 2015). Our results show that the ΔCt for one stored mRNA is highly correlated with seed aging time. Thus, the changes in stored mRNAs (represented by ΔCt values) can serve as a more precise method of assessing seed aging. In addition to the qPCR-based methods, long-read DNA sequencing technologies (Lu *et al*., 2016; Oikonomopoulos *et al*., 2016) may also be used for determining the changes in transcript length. However, analyzing one or a few mRNAs is technically much simpler and requires less effort, cost and time than sequencing entire cDNA libraries. Further, multiple steps of DNA manipulation including rounds of DNA purification are needed for cDNA library preparation, for instance for MinION sequencing (https://community.nanoporetech.com/protocols). The fragment size distribution in the resulting cDNA libraries may not truly represent the fragment size distribution in the initial RNA samples due to the differences in the binding efficiency to the DNA purifying matrix among different DNA fragments, affecting the quantitative analysis based on the cDNA fragment lengths.

Since seeds age with different speeds under different conditions, seed storage time may not be a good indicator of the seed aging status. Stored mRNAs can serve as more reliable biomarkers. We observed that the AA and NA seeds with the same seed viability (0.5% germination percentage) showed similar ΔCt values for the fragments on B16 and B20 (Figure 3.9) despite the huge differences in their aging time between AA (15 days) and NA (17 yrs) seeds. Thus, one possible way of measuring the aging status of different NA seeds is to establish a reference aging timeline using AA seeds, and then map the NA seed samples to a point on the reference timeline, in a way similar to the use of a reference protein to

determine the concentrations of protein samples.

In this study, we developed new methods to quantify the changes in seed stored mRNAs and estimate the rate of mRNA breakdown at the one nucleotide level. These methods should facilitate the studies of seed stored mRNAs in plants. The β value or frequency of breaks per nucleotide per day should make it easier to quantify the effects of different conditions on stored mRNAs and seed aging. Furthermore, these methods should also be applicable for analyzing RNA degradation in other plants and non-plant systems.

4. CHAPTER FOUR – SEED STORED mRNA DEGRADATION IN WHEAT AND CANOLA SEEDS

4.1 Introduction

The findings on stored mRNAs in Arabidopsis seeds are interesting. It is an important question whether similar changes occur in crop seeds, and whether stored mRNAs can be used as markers to assess the seed aging of a large number of seed germplasm materials. Also, different plant species differ in genome complexity and seed longevity or aging speed. Stored mRNAs may degrade differently in different plants and a complex genome may make it more difficult for the analysis by qPCR due to the presence of related genes and sequences. Thus, it is important to determine how seed stored mRNAs change during seed aging in crop species with a more complex genome. Wheat and canola are two of the most important crops in Canada. In this chapter, mRNA degradation will be analyzed in the seeds of common wheat (*T. aestivum*) and canola (*B. napus*) to validate that (1) whether mRNA degradation during seed aging is a general phenomenon; and (2) whether the ΔCt could be used as an accurate and variable biomarker and be applied on crop species.

4.2 Seed stored mRNA degradation in wheat seeds

4.2.1 Comparison of wheat germination assays

Two germination assays were compared in order to find a reliable and simple method for wheat germination. One method was the traditional wheat germination method serving mainly to assess seed quality for agricultural and field applications. It was routinely used in the AAFC-Plant Gene Resource of Canada (AAFC-PGRC) to test seed germination for some species. The other method was the "plate" method used for Arabidopsis germination.

In the traditional method, the seeds were distributed evenly between two layers of wetted germination paper. The papers were rolled (to facilitate handling) and placed into a plastic bag. After 7 days incubation (20 ^{o}C with 8 h light and 16 h dark), seed germination was examined and the germination percentage was obtained. For the method adopted from the "plate" method, wheat seeds were sterilized and plated on plates with ½ MS (Murashige

and Skoog, 1962) medium containing 1% sucrose and 0.7% agar. After 7 days of incubation, the germination percentage was obtained. As shown in Figure 4.1, while the seeds in the classical method did not germinate evenly and also microbial growth on the paper was apparent (Figure 4.1A), the sterilized seeds germinated evenly in the Petri plates (Figure 4.1B).

The classical method is simple and economical, requiring only a small piece of paper material. Thus, it has been widely used for assessing seed quality and agricultural applications. However, one major problem with the method is the growth of microbes which can affect the consistency and accuracy of the germination estimate, which is avoided through germinating sterilized seeds on medium. Since we aimed to determine the correlation between seed germination percentage and changes in stored mRNAs, a method with better accuracy is preferred. Thus, the medium-plate method was used.

4.2.2 Wheat accelerated aging assay

AA treatment is needed for this study since NA takes many years. Several AA assays have been reported and used, however, most of them use rather harsh conditions, for instance, high temperatures (45 ^{o}C or above) (Ganguli and Senmandi, 1993; Krishnan *et al*., 2004; Vanessa Ocom Menezes, 2014). Since proteins and enzymes are easily denatured and cellular structure disrupted at such high temperatures, seed aging under these conditions may not reflect what happens during the process of natural seed aging under long-term storage. We thus aimed first to establish an accelerated aging scheme at a lower temperature which could reflect the natural seed aging. I used a protocol similar to the one we used for Arabidopsis seed aging treatments, except that a higher temperature (40 ºC) was used.

The results for wheat AA treatment are shown in Figure 4.2. First, the germination percentage curve of the AA wheat seeds was similar to that of the AA Arabidopsis seeds. For the first 9 days under the seed aging conditions used, there was only a small change in germination percentage. After that, seed germination percentage decreased from above 80% to about 10%, and completely lost the ability after 24 days of AA treatment. Second, wheat seeds took longer time to age, or had a longer “lifespan” than Arabidopsis seeds. It took

wheat seeds 24 days of AA treatment at 40 ºC to completely lose the ability to germinate compared to 16 days for Arabidopsis seeds at 37 ºC.

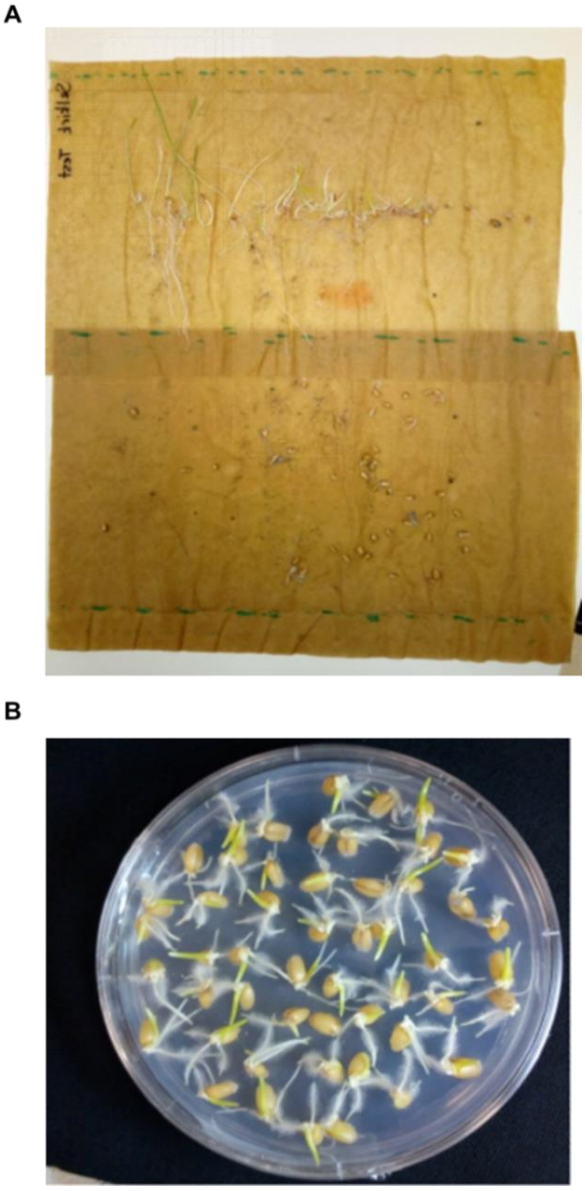


Figure 4.1 Comparison of two wheat seed germination assays. (A) Seeds were germinated using the traditional method. One hundred seeds were distributed in two layers of germination paper, which were moistened with distilled water in a volume equivalent to 2.5 times the mass of the dry substrate. The paper was rolled up and kept at 20 ^{o}C in the seed incubator. Germination assessments were performed on the seventh day after sowing. (B) Germination assay using the "plate" method. Seeds were sterilized with 20% bleach for 15 min, followed by four rinses with sterilized ddH_2O. Fifty seeds were plated on ½ MS plates, and the seed germination percentage was assessed after ten days of incubation in a tissue culture chamber at 20 ^{o}C.

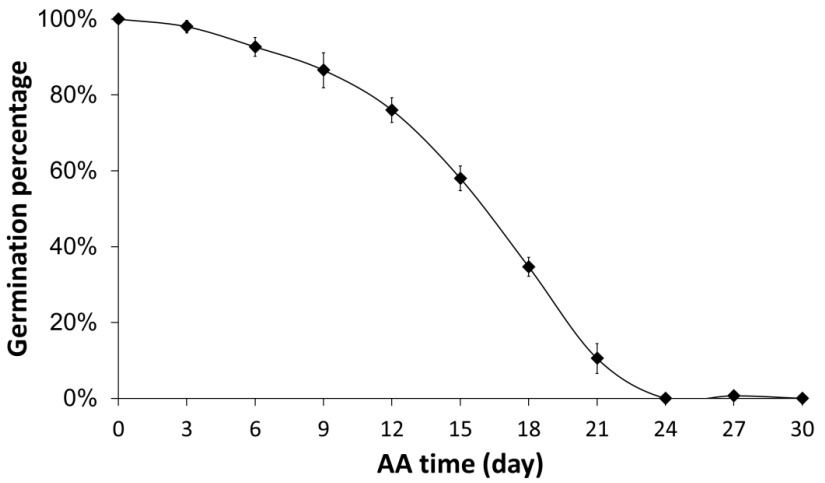


Figure 4.2 Change in germination percentage for wheat seeds aged in an accelerated aging assay. Seeds were treated at 40 ºC and under 82% humidity for the indicated times (one unaged control and ten AA seed samples). For each seed lot, seed germination test was conducted with three replicate plates with 100 seeds each and the standard errors are shown.

4.2.3 RT-PCR analysis of stored mRNAs during wheat seed aging

To study the mRNA changes in aged seeds, 19 candidate wheat genes were surveyed for the presence of stored mRNAs and for simplicity and ease of reference they were referred to by the codes such as W1 ("W" standing for wheat), W2, etc. (Table 2.3). Among the 19 genes, W1-W8 were selected randomly, and W9-W13 were wheat homologs for five Arabidopsis genes (B10, B14, B16, B20, and C22 respectively) used in the Arabidopsis analysis. In addition, since common wheat is a hexaploid plant and the majority of genes would have multiple related sequences, I also identified and used several wheat genes (W14 – W19) which appear to be unique based on sequence analysis. For the initial survey by RT-PCR, a fragment size about 950 bp (starting from or near the STOP codon) was used. The results showed that under the present RT-PCR conditions a clear and specific cDNA band was amplified for nine of the nineteen genes (W2, W3, W7, W10, W11, W12, W14, W16, and W17) (Figure 4.3). W2, W3, W10 and W12 had a prominent and specific band, and thus were used in further analyses. W7 was not used in further analyses because its coding sequence (open reading frame) is less than 900 bp.

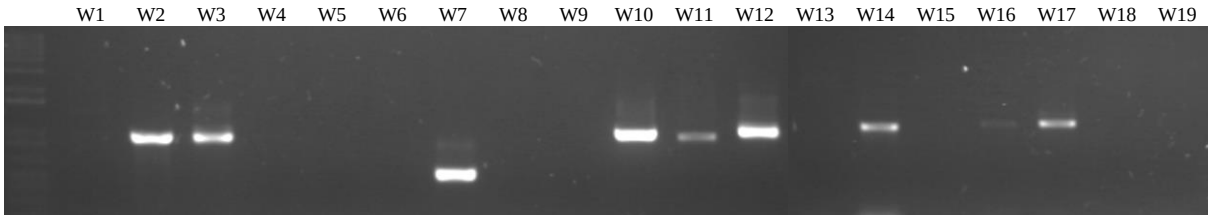


Figure 4.3 Results of RT-PCR analysis of 19 genes to survey the presence of stored mRNAs in wheat dry seeds. Nineteen wheat genes (as listed in Table 2.3) were used to determine the presence of stored mRNAs by RT-PCR. cDNA reverse-transcribed from total RNA of unaged dry seeds was used in PCR with primers specific for each of the genes. The PCR products were subjected to electrophoresis in 1% agarose gel.

4.2.4 qPCR analysis of stored mRNAs in aged wheat seeds

The data from Arabidopsis show that the ΔCt value for a stored mRNA between aged and unaged seeds increases with the transcript length being analyzed (Figure 3.9). To determine whether this is true in wheat, the relationship between ΔCt value and fragment length in 30-day aged wheat seeds was determined using W2. For sequence specificity and to avoid a high GC content, primer positions might slide slightly from a predetermined position on the template and the lengths between the forward and reverse primers might not be strictly in hundreds. Using wheat control seeds and seeds aged for 30 days, the levels of the six fragments on gene W2 ($W2_{107bp}$, $W2_{295bp}$, $W2_{490bp}$, $W2_{703bp}$, $W2_{933bp}$, and $W2_{1131bp}$, see Table 2.3) were analyzed by qPCR. The resulting ΔCt values for the six fragments were plotted against the fragment sizes. As shown in Figure 4.4, there was a very strong correlation between ΔCt and fragment size ($R^2 = 0.9959$ and $P < 0.0001$).

To analyze the relationship between ΔCt value and aging time for wheat, we used a set of AA seeds treated under the same accelerated aging conditions. Since it has been observed in Arabidopsis that the regression slope of the linear correlation between ΔCt value and aging time increases with fragment length, I determined the ΔCt values for different fragments of W2 (they are $W2_{107bp}$, $W2_{295bp}$, $W2_{490bp}$, $W2_{703bp}$, $W2_{933bp}$, and $W2_{1131bp}$ respectively, see Table 2.3). The ΔCt values for each fragment were generated and plotted against seed aging time. As shown in Figure 4.5A, the correlation coefficients for all fragments were greater than 0.90, and the slopes of the ΔCt regression lines increased gradually with the fragment length, similar to what had been observed in Arabidopsis.

To minimize the variance caused by cDNA concentration among samples, the ΔCt values for longer fragments can be normalized with the values of a short fragment as introduced in Chapter 3. In wheat, ΔCt values for $W2_{295bp}$, $W2_{490bp}$, $W2_{703bp}$, $W2_{933bp}$, and $W2_{1131bp}$ were normalized by the values for $W2_{107bp}$, and the normalized ΔCt (or ΔΔCt) values were plotted against seed aging time as shown in Figure 4.5B. The normalization improved the correlations for all fragments, showing that normalizing the ΔCt values using a much shorter fragment reduced data variation. However, the correlations using ΔCt value without normalization were already very strong and normalizing the ΔCt value has a limited impact on the correlations. Thus, when the correlations are very high (e.g. with an $R^2 > 0.95$ and $P < 0.001$) and since the gain by the normalization is relatively small, the ΔCt value can be effective without the need for normalization.

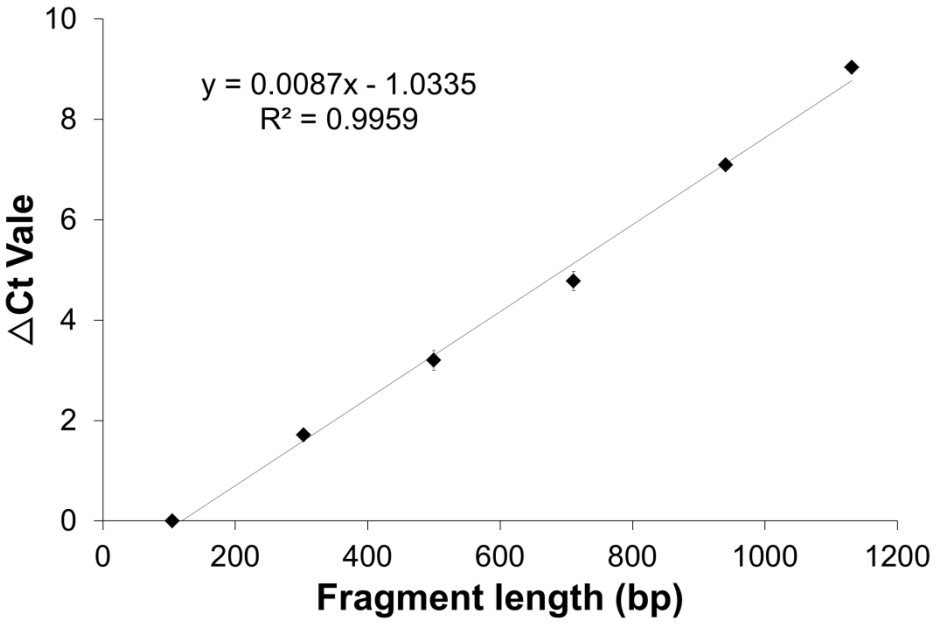
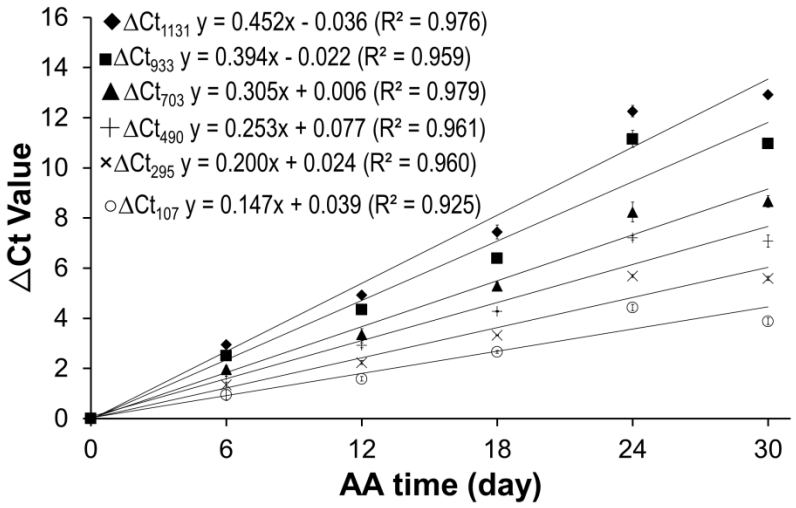


Figure 4.4 The relationship between ΔCt value and cDNA fragment size analyzed using one wheat gene. cDNAs of the control (100% germination) and 30-day AA seeds were used to obtain ΔCt values between the aged and control samples for different fragments of the W2 gene (i.e. $W2_{107bp}$, $W2_{295bp}$, $W2_{490bp}$, $W2_{703bp}$, $W2_{933bp}$, and $W2_{1131bp}$ respectively, from the -59 position of the STOP codon). The ΔCt values are plotted against the fragment sizes. Each ΔCt value is based on three technical repeats with the standard error bar shown. The linear regression was obtained with Microsoft Excel 2010 and has a $P < 0.0001$.

A



B

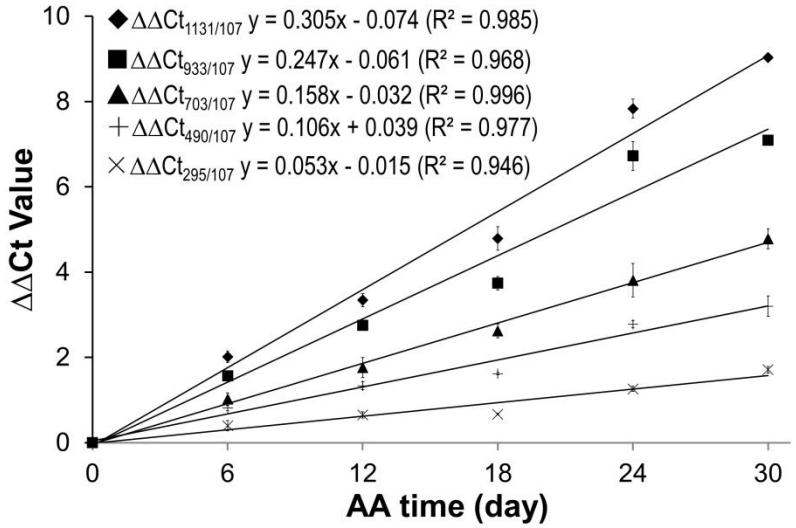


Figure 4.5 Correlation of mRNA degradation with seed aging time analyzed using different cDNA fragment lengths of gene W2. cDNAs of the control and five acceleratedly aged (AA) seeds for the indicated aging time (days) were used. Six different fragments of the gene W2 were used. Each ΔCt value is based on three technical repeats with the standard error bar shown. (A) Correlation between ΔCt values for six different fragments of W2 and aging time. The subscript after ΔCt indicates the cDNA fragment length (as ΔCt_{1131bp} indicates the ΔCt is for the $W2_{1311bp}$ fragment). (B) Correlation between normalized ΔCt values for five different fragments of W2 and aging time. The normalized ΔCt or $\Delta\Delta_{Ct1131/107bp} = \Delta Ct_{1131bp} - \Delta Ct_{107bp}$. Linear regression functions were calculated with Excel 2010 and all have a $P < 0.01$.

I then determined how stored mRNAs of different genes changes during wheat seed aging. One issue that needs consideration is the template abundance since if a stored mRNA has low abundance, the level in the aged seeds could be more difficult for qPCR analysis and cause more variation. Thus, we selected four wheat genes (W2, W3, W10 and W12) that showed high and similar abundance of stored mRNAs based on the RT-PCR results (Figure 4.3) and a similar fragment length around 950 bp was analyzed. As shown in Figure 4.6, the ΔCt value showed a high correlation with aging time with the correlation coefficient near or above 0.90 for all four genes. Further, the slopes of the ΔCt regression lines are very similar among them, suggesting that these fragments have a very similar mRNA degradation rate.

The β values for the four wheat genes (stored mRNAs) were obtained from the ΔCt values. Since the correlations between the ΔCt value and aging time were high (with R^2 values for one gene > 0.95 and the fourth one close to 0.90), the ΔCt values were used without normalization (with the ΔCt values of a short fragment of the same stored mRNA). The β values, as shown in Table 4.1, were for the most part similar, with the value for the gene W2 being slightly higher at 2.94 x 10^{-4}. The overall average β value for the four wheat genes was 2.26 x 10^{-4}. In comparison, the overall average β value for Arabidopsis genes was 0.98 x 10^{-4} (Table 3.7). The higher value for the wheat seeds aged at 40 ºC compared to the Arabidopsis seeds aged at 37 ºC is consistent and supports the observation made in Arabidopsis that the β value increases with increasing aging temperature (Table 3.7 and Figure 3.16).

Table 4.1 Estimated β values for different wheat fragments at different aging times.

Gene code	Genes[1]	Fragment size (bp)	β value (break per nucleotide per day x 10^{-4})[2]					
			Day 6	Day 12	Day 18	Day 24	Average	Overall
W2	$W2_{950bp}$	950	3.30	2.68	2.62	3.14	2.94±0.15	2.26±0.12
	$W3_{950bp}$	950	1.86	1.49	1.71	2.46	1.88±0.18	
	$W10_{955bp}$	955	2.12	1.94	2.15	2.42	2.16±0.09	
	$W12_{950bp}$	950	1.80	1.99	2.41	2.07	2.07±0.11	
Average			2.27±0.30	2.03±0.21	2.22±0.17	2.52±0.19		

The ΔCt values between the control and aged seeds as shown in Figure 4.6 to calculate the β values using the equation (8) in the Materials and Methods. Each β value is based on ΔCt values of three technical replicates and the standard error is indicated.

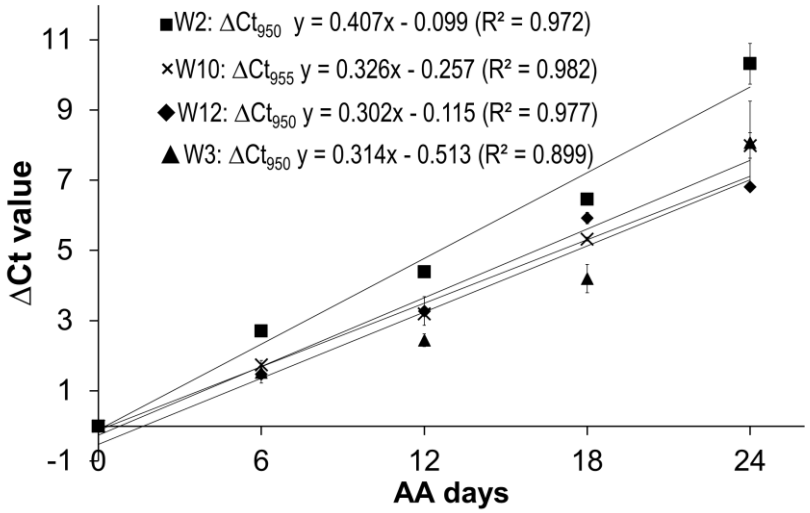


Figure 4.6 The relationship between the ΔCt values of stored mRNAs of four wheat genes and seed aging time. cDNAs of control and four acceleratedly aged (AA) seeds for the indicated aging time (days) were used. Stored mRNAs of four genes (with the fragments of $W2_{950bp}$, $W3_{950bp}$, $W10_{955bp}$, and $W12_{950bp}$) were analyzed. Each ΔCt value is based on three technical repeats, with the standard error bar shown. The linear regressions were produced with Excel software. The x variable for W2, W10 and W12 in the linear regression functions has a $P < 0.0001$, while W3 has a $P < 0.01$ in the *F*-test.

4.3 Seed stored mRNA degradation in *B. napus* seeds

4.3.1 Decrease of seed viability with aging time for *B. napus*

Since *B. napus* seeds take years to age under cold temperature (4 °C) condition, AA seeds were used. For the aging treatments, the same AA conditions used for Arabidopsis seeds (37 °C and 83% RH) were used for *B. napus* seeds. After the seed aging treatments, AA seeds were collected and their germination percentages were determined. As shown in Figure 4.7, the change in seed germination percentage over the aging time had a similar trend to what was observed for Arabidopsis (Figure 3.11). However, *B. napus* seeds took longer time to lose their viability completely (27 days for *B. napus* seeds vs 16 days for Arabidopsis seeds, see Figure 3.11). In addition, under similar aging conditions, the overall decreasing curve in the germination percentage for *B. napus* is not as sharp as for Arabidopsis.

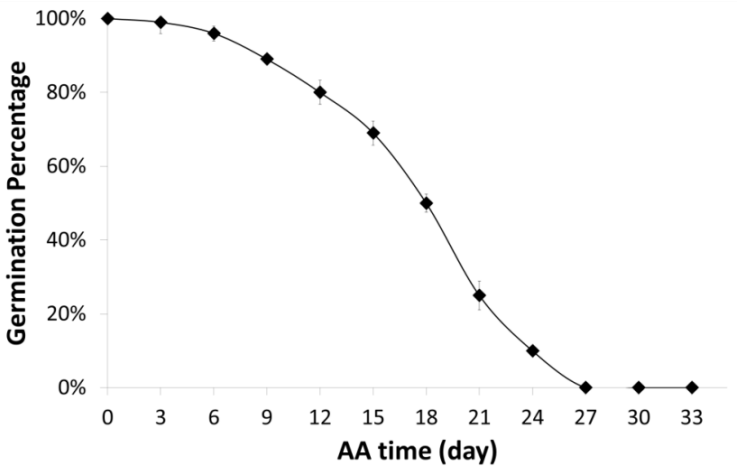


Figure 4.7 The change in germination percentage in *B. napus* seeds with AA days. Seeds were treated with AA conditions for the 11 indicated times (at 37 ºC and 83% humidity). The germination percentages were calculated based on three plates (with 100 seeds in each plate) and the standard errors are shown.

4.3.2 Survey of the presence of stored mRNAs in *B. napus* seeds

Nineteen randomly selected candidate genes were surveyed firstly for the presence of stored mRNA in *B. napus* seeds. For simplicity and ease of reference, these genes were referred to by the simple codes such as Bn1 (*B. napus* 1) (Table 2.4). Also, since mRNA degradation is determined mainly by the length of transcript analyzed, the fragment length is indicated as the subscript of the code when needed, such as $Bn1_{1500bp}$, $Bn2_{2000bp}$, and $Bn2_{1100bp}$. The initial survey by RT-PCR revealed that at least ten genes (Bn1, Bn2, Bn3, Bn4, Bn9, Bn11, Bn12, Bn13, Bn17, and Bn18) showed a clear and specific band and five genes (Bn7, Bn8, Bn14, Bn16, and Bn19) showed a weak but visible band on the agarose gel (Figure 4.8).

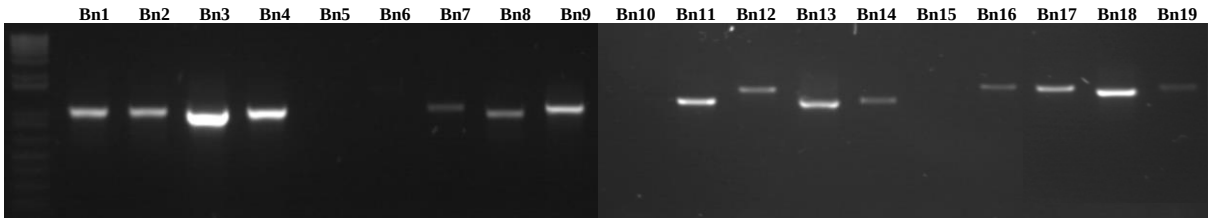


Figure 4.8 The presence of stored mRNAs in *B. napus* dry seeds for the genes listed in Table 2.4. cDNA reverse-transcribed from total RNA of unaged dry seeds was used in PCR with primers specific for each of the genes. Totally 19 *B.napus* genes were surveyed. The PCR products were subjected to electrophoresis in 1% agarose gel.

4.3.3 Quantitative analysis of seed stored mRNA in aged *B. napus* seeds

Experiments were then conducted to analyze the relationships between mRNA decrease, stored mRNA (or cDNA) fragment size, and seed aging time in *B. napus* similar to what was done in Arabidopsis and wheat. Firstly, the correlation between ΔCt value and stored mRNA fragment length was determined. For different fragments on the candidate gene, due to the consideration for selecting a proper sequence for a primer, some of the forward primers were located at a distance from the reverse primer not in exact hundreds of basepairs e.g. 1000 bp. Using *B. napus* control seeds and seeds aged for 30 days, the levels of the five fragments on the gene Bn12 ($Bn12_{148bp}$, $Bn12_{487bp}$, $Bn12_{850bp}$, $Bn12_{1169bp}$, and $Bn12_{1500bp}$ in Table 2.4) were analyzed by qPCR. As shown in Figure 4.9, there was a tight correlation between ΔCt value and fragment size ($R^2 = 0.949$ and $P < 0.01$), indicating that the level of undamaged stored mRNAs decreased with the increasing length of the cDNA fragment analyzed, consistent with what was observed in Arabidopsis and wheat.

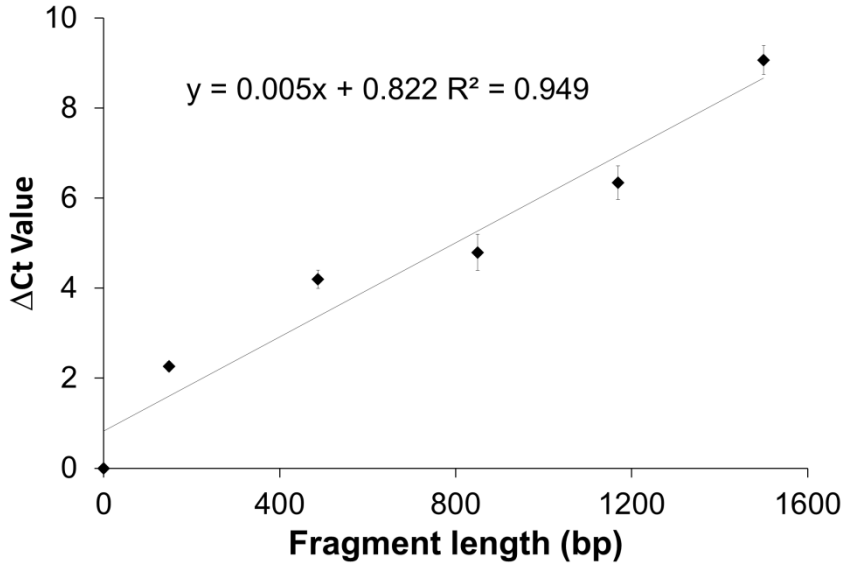


Figure 4.9 The relationship between ΔCt value and cDNA fragment size analyzed using *B. napus* gene Bn12. cDNAs of the control (100% germination) and 30-day AA seeds were used to analyze different fragments of the Bn12 gene; and the fragments are $Bn12_{148bp}$, $Bn12_{487bp}$, $Bn12_{850bp}$, $Bn12_{1169bp}$, and $Bn12_{1500bp}$ respectively. Each ΔCt value is based on three technical repeats with standard error bar shown. The linear regression was obtained with Microsoft Excel 2010 and had a $P < 0.01$.

Since the ΔCt value between the aged and control *B. napus* seeds increased linearly with the fragment length of Bn12, the correlation between the ΔCt value and seed aging time was further determined for each of the five fragments using a set of seeds aged for different lengths of time (0, 6, 12, 18, 24, and 30 days). As shown in Figure 4.10, there were good correlations between the ΔCt value and seed aging time for each of the five fragments on gene Bn12. In addition, the slopes of the regression lines increased gradually with fragment length in a similar way to what was observed in Arabidopsis and wheat.

Next, stored mRNA degradation was determined and compared for different genes which have a similar fragment length. Thus, $Bn11_{1490bp}$, $Bn12_{1500bp}$ $Bn13_{1486bp}$, $Bn14_{1500bp}$, and $Bn17_{1500bp}$ were analyzed using a set of aged seeds and the resulting ΔCt values were plotted against seed aging time. As shown in Figure 4.11, the slopes of the regression functions are very similar for the five genes, suggesting a similar rate of mRNA degradation. These results suggest that stored mRNAs are degraded similarly in *B. napus* compared to Arabidopsis and wheat, and the level of a stored mRNA can be used as a molecular marker for seed aging in *B. napus*.

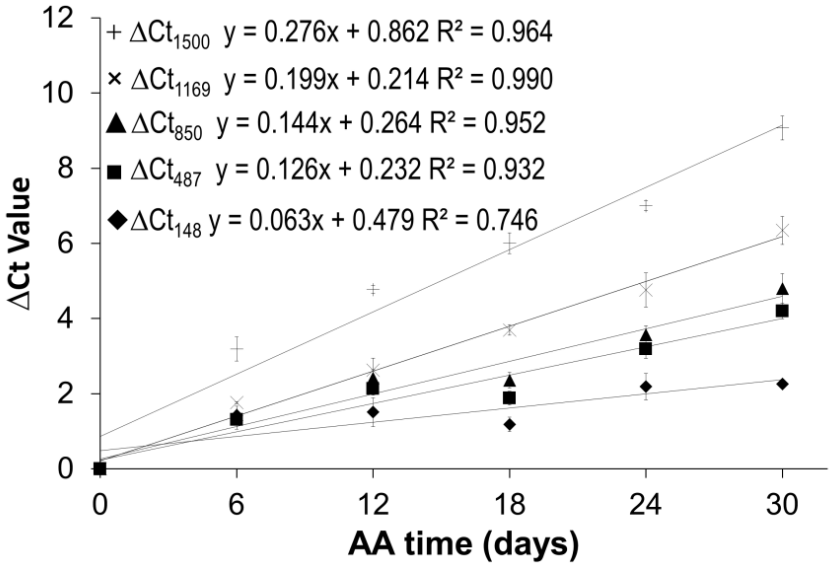


Figure 4.10 Correlation of mRNA degradation with seed aging time analyzed using different cDNA fragments of Bn12. cDNAs of the control and five AA seed lots for the indicated aging time (days) were used. Different fragments of the gene Bn12 were used: 148 bp, 478 bp, 850 bp, 1169 bp, and 1500 bp. Each ΔCt value is based on three technical replicates with the standard error bar shown. The *P* values of the regression lines are less than 0.05 for $Bn12_{148bp}$ and less than 0.001 for $Bn12_{478bp}$, $Bn12_{850bp}$, $Bn12_{1169bp}$ and $Bn12_{1500bp}$ in the *F*-test.

From the ΔCt values, the β values for the five *B. napus* genes were generated. As shown in Table 4.2, the β values for the five *B. napus* genes were quite similar, confirming that different stored mRNAs are degraded at a constant rate over the aging time as observed in Arabidopsis (Table 3.7) and wheat (Table 4.1). Interestingly, the *B. napus* genes have a similar overall β value with that of the Arabidopsis genes (1.37±0.05 in Table 4.2 VS 1.25±0.07 in Table 3.7 and 0.98±0.01 in Table 3.7) when the seeds were aged under 37 ºC and 83% RH, suggesting that the stored mRNAs are degraded with comparable rates in seeds of the two plant species.

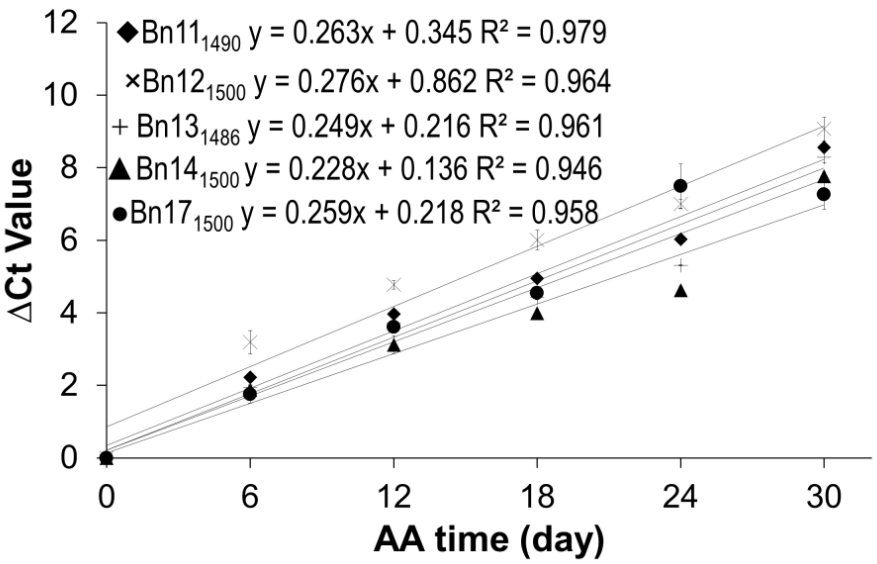


Figure 4.11 The relationship between the ΔCt values of stored mRNAs of six *B. napus* genes and seed aging time. cDNAs of control and five acceleratedly aged (AA) seed lots for the indicated aging time (days) were used. The ΔCt values between the aged and control samples for Bn11$_{1490bp}$, Bn12$_{1500bp}$, Bn13$_{1486bp}$, Bn14$_{1500bp}$, and Bn17$_{1500bp}$ (the fragment length analyzed is indicated by the subscript) were determined. Each ΔCt value is based on three technical replicates with the standard error bar shown. The linear regressions were produced with Excel 2010 and every x variable in the linear regression functions had a $P <$ 0.01.

Table 4.2 Calculated β values for different *B. napus* fragments at different aging times.

Fragment[(1)]	β value (break per nucleotide per day x 10^{-4})						Overall
	Day 6	Day 12	Day 18	Day 24	Day 30	Average	
$Bn11_{1490bp}$	1.73±0.04	1.54±0.06	1.28±0.04	1.17±0.04	1.33±0.06	1.41±0.05	
$Bn12_{1500bp}$	2.46±0.24	1.84±0.04	1.54±0.07	1.35±0.03	1.4±0.05	1.72±0.09	
$Bn13_{1486bp}$	1.51±0.05	1.44±0.02	1.15±0.02	1.03±0.00	1.29±0.02	1.28±0.02	1.37±0.05
$Bn14_{1500bp}$	1.44±0.10	1.20±0.10	1.02±0.02	0.89±0.02	1.20±0.02	1.15±0.05	
$Bn17_{1500bp}$	1.35±0.20	1.39±0.01	1.17±0.03	1.44±0.01	1.12±0.06	1.29±0.06	
Average	1.70±0.13	1.48±0.05	1.23±0.04	1.18±0.02	1.27±0.04		

The β value values were calculated using the ΔCt values from qPCR analysis, as shown in Figure 4.11. Each β value is based on ΔCt values of three technical replicates, and the standard error is indicated.

(1) The fragments analyzed are indicated. For instance, $Bn11_{1490bp}$ indicates that a 1490 bp fragment of Bn11was analyzed.

4.4 Discussion

4.4.1 New methods are useful for quantifying stored mRNA degradation

The decrease of RNA integrity in aged seeds has been reported in many plant species. However, it is difficult to quantify the changes in RNA integrity based on isolated total RNA (Boniecka *et al*., 2019), or smaller RNA bands on the electrophoresis graphs (Brocklehurst and Fraser, 1980), and particularly the relationship between RNA integrity seed aging time. Recently, the RNA integrity number (RIN), a parameter generated using Agilent Bioanalyzer was found to decrease in aged soybean seeds and correlate with seed aging time (Fleming *et al*., 2017). However, since the RIN is an integrated parameter based on multiple features of a RNA electrophoresis graphs, it is affected not just by RNA degradation, but also the purity, concentration and abnormal peaks. Our analysis showed that the RIN values were similar for the Arabidopsis RNA samples from seeds aged for different times (Table 3.2).

Stored mRNA degradation has also been reported recently by RNA-seq analysis using nanopore (MinION) DNA sequencing (Fleming *et al*., 2018b). However, to the best of our knowledge, currently there is no method available that could determine the rate of stored

mRNA degradation based on the large-scale sequencing data. Thus it is difficult to use the sequencing data to study the dynamics of stored mRNA degradation over seed aging time. Further, as discussed in the last Chapter, due to multiple steps of DNA manipulation particularly DNA fragment purification, the length representation of different DNA fragments in the cDNA libraries might be altered from that in the original RNA samples. In addition, large-scale sequencing analysis is much more costly and requires greater efforts than the analysis of a limited number of stored mRNAs by qPCR.

In the current study, through analyzing stored mRNA degradation in Arabidopsis (Chapter 3), wheat and canola (this chapter), we observed very tight correlations between the ΔCt value and seed aging time in the three different species. Typically the correlation coefficients have an R^2 value of more than 0.95 for longer fragments. This correlation relationship is valid from the unaged seeds to an aged time far beyond the point at which aged seeds lose the ability to germinate completely. Comparing the high correlation coefficients observed between the ΔCt value and seed aging time in our study, the average correlation coefficient between RIN and seed aging time was around 0.6 (Fleming *et al*., 2018b); and no quantitative correlation can be developed with the lowered total RNA amount (Boniecka *et al*., 2019) and the smaller RNA bands on the electrophoresis graphs (Brocklehurst and Fraser, 1980) in aged seeds. Regarding other parameters such as seed germination, relative root length, relative fresh weight and electrolyte conductivity, our analysis on Arabidopsis seeds showed that those parameters either have a non-linear correlation pattern or a low coefficient with seed aging time. Thus, the ΔCt value based on the level of stored mRNAs (with one stored mRNA being sufficient) generally has a better correlation with seed aging time than all other known methods and can be potentially a more precise means for assessing seed aging status.

The tight correlations of ΔCt with mRNA fragment length and seed aging time, observed consistently in three different plants, strongly suggest that the damage on a stored mRNA (1) occur randomly along the length of a stored mRNA and (2) accumulate evenly with seed aging time under the same conditions (See discussion in the last chapter). Based on these characteristics of stored mRNA degradation during seed aging, we have developed a

new parameter, the β value representing the average frequency of "breaks" per nucleotide per day. The β value can be estimated based on the ΔCt values using an equation we developed. Although it is an indirect estimate based on the level of cDNA derived from the stored mRNA, to our knowledge, currently it is the only parameter or method that allows analysis of the rate of degradation of a stored mRNA.

4.4.2 Comparisons of stored mRNA degradation and seed aging in three different plant species raise some interesting questions

Traditionally, the differences among plant species in seed aging have been compared and evaluated using seed germination. However, the readout from a seed germinaiton assay is qualitative in nature. For one seed, the possible outcome is either yes (a seed can germinate) or no (the seed cannot germinate), while when a sample of seeds are analyzed the germination percentage does not follow a strict linear relationship with seed aging time, as shown by others and my results here. Further, the seed germination percentage reveals little about the changes occuring to macromolecules during seed aging.

The β value established in this study can allow quantitative and more precise comparisons. The results indicate that the β values for Arabidopsis were consistent for different fragments (Table 3.6) and for different stored mRNAs genes (Table 3.7) throughout the aging process under the same conditions. Further analysis suggests that the β value increased with increasing temperature in Arabidopsis aging (Table 3.7). Thus, the β value can be very useful for comparing stored mRNA degradation and seed aging (since the degradation of stored mRNAs has a very tight correlation with seed aging time) under different aging conditions and among different species.

One interesting question is whether stored mRNAs in different seeds are degraded with similar or different rates. In the current study, Arabidopsis and *B. napus* seeds were aged under the same conditions (37 ºC and about 82-83% RH). The β values for six Arabidopsis and five *B. napus* stored mRNAs (Table 3.7 vs Table 4.2) were comparable, with the total average being 0.98×10^{-4} for the Arabidopsis stored mRNAs and 1.36×10^{-4} for the *B. napus* stored mRNAs. For the wheat seeds, since they age slower and retain the ability to

germinate for a longer time than the Arabidopsis seeds under the same accelerated aging conditions, 40 ºC was used for the accelerated aging. At 37 ºC (and 82-83% RH), the Arabidopsis seeds took about 16 days to completely lose the ability to germinate while the wheat seeds took 24 days at 40 ºC. The β value for Arabidopsis B16 at 40 ºC was 2.2 x 10^{-4} (Table 3.7). In comparison, the overall average β value for four wheat genes when aged at 40 ºC was 2.26 x 10^{-4} (Table 4.1). Considering that the seed aging treatments and RNA isolation for different plants were performed at different times for three plant species and further possible differences in seed quality and water content in the dry seeds prior to the treatments, the β values for the three plants are remarkablly similar. This is a very interesting observation. Although awaiting further confirmation with the analysis of more stored mRNAs, it suggests that stored mRNAs for the three different plants are degraded with comparable rates under the same conditions.

These results also raise an interesting question regarding the differences in seed longevity among different plants. Our data show that it took 16 days and 27 days for Arabidopsis and *B. napus* seeds, respectively, to completely lose the ability to germinate when aged at 37 ºC under the same conditions, while it required about 24 days and a higher aging temperature of 40 ºC for wheat seeds to do so. If the stored mRNAs are degraded at similar rates in seeds of the three different plants, why do the seeds have different times of longevity when aged under the same conditions? There are several possibilities. First, the total amounts of the stored mRNAs may be very different in seeds of different plants. The larger wheat and *B. napus* seeds may have much higher total amounts of stored mRNAs and they could last for longer times during the seed aging process. Second, the common wheat is a hexaploid and *B. napus* is a tetraploid, while Arabidopsis is a diploid. In the polyploids, there are more related genes in a gene family. As a result, there is a higher level of functional redundancy among the stored mRNAs in wheat and *B. napus* seeds compared to the Arabidopsis seeds. Third, seed germination depends on many different macromolecules and cellular structures to be functional. Seeds of different plants may have a different threshold requirement for the minimal amount of stored mRNAs at the early phase of seed germination. Our method that allows quantitative analyses of the level and rate of degradation of stored

mRNAs make it possible to address some of the questions in the future.

5. CHAPTER FIVE - GENERAL DISCUSSION AND FUTURE WORK

5.1 General discussion

5.1.1 Importance and current status of seed aging research

Plant genetic resources are essential in providing food, fiber, and feed, but are being lost increasingly (FAO, 2010). It has been reported recently that plants are becoming extinct at a much faster and alarming rate for the modern time after 1900 AD than the previous time before 1900 (Humphreys *et al*., 2019). *Ex situ* storage of the 7.4 million seed accessions is one of the most effective ways to conserve plant genetic resources (FAO, 2010). However, seeds lose their germination ability and eventually die during storage (Ventura *et al*., 2012). Performing seed viability tests and regenerating stored seed stocks for huge numbers of accessions in the seed banks pose practical challenges. Therefore, effective and accurate methods in assessing seed deterioration over time are critical in maximizing seed storage duration in seed banks (Walters *et al*., 2005 ; Hay and Probert, 2013 ; van Treuren *et al*., 2013).

At the same time, seed quality is also central to agricultural production, and has been regarded as an essential agronomic trait. Most of the crops start with the seedling establishment in the field, and uniform seedling establishment sometimes determines the final yield (Finch-Savage and Bassel, 2015). However, before being sown in the field, seeds are transported and stored in seed companies for different durations of time. Sometimes, inappropriate storage conditions and long storage times may undermine seed vigor, resulting in sub-optimal performance of seedling establishment in the field such as delayed seed germination and lowered plant density (Boniecka *et al*., 2019; Gray, 1976; Salter *et al*., 2009; Wurr, 1983). The delayed seed germination may shorten the growing season and complicate the crop management (David, 1987). To avoid unnecessary losses, quick and simple assays are required to check seed vigor before launching commercial seeds into the market (Boniecka *et al*., 2019).

However, all seeds deteriorate gradually during storage, leading to the loss of seed vigor and eventually seed viability (Walters, 2008). Despite much research efforts, the

understanding of the molecular basis of seed aging is still limited. The fact that different plants have different longevity patterns and aging speeds implies that seed aging is not simply due to the exhaustion of energy. In general, seed aging is regarded as the result of gradual accumulation of deteriorating molecules (e.g. ROS and cyanide) as well as molecular damage on macromolecules (e.g. DNA, RNA, protein, and lipids), resulting in dysfunctional molecules and cells (Walters, 2008). In addition, the enzymatic repair systems are not functional under dry state under storage conditions, resulting in irreversible damage which may further hamper the turnover of macromolecules during seed germination (Groot *et al*., 2015).

Unfortunately, there are no perfect tools to evaluate seed aging status in managing seed banks or commercial seeds. The most common method of determining seed viability and vigor is seed germination under defined laboratory conditions. However, seed germination is based on a singular break-point event that a seed is no longer able to germinate. It reveals relatively little about the changes occurring to the seed during seed storage or seed aging before this point. Therefore, a method relying on the internal changes occurring to a seed, which are taking place continuously over the seed aging time and can be easily measured, would be better for assessing seed aging status. There have been various methods based on certain aspects of seed for viability analysis such as seedling growth, tetrazolium test, electrical test, and ROS scavenging ability (Fu *et al*., 2015). However, most of these assays are less than desirable in terms of accuracy, sensitivity and strong quantitative relationship with seed aging status. More accurate and reliable methods based on quantifiable cellular and molecular changes will be useful for seed aging research and seed stock management in seed banks.

5.1.2. Novel findings have been made and methods developed on stored mRNA degradation during seed aging in three different plants

In this research project, we focused on seed stored mRNAs and their changes during seed aging, since stored mRNAs are considered to be required for the early phase of seed germination and at the same time they are prone to degradation. Several major findings have

been made that have important implications for seed aging research and seed viability assessment. First, we developed for the first time quantitative methods (and mathematical equations) to estimate the changes of a stored mRNA at the full length (or fragment) level and one nucleotide level. Second, several fundamental characteristics of stored mRNA degradation were discovered, supported by the results from the model Arabidopsis as well as from canola and wheat. Third, the present findings challenge some traditional concepts for explaining seed aging. Finally, our results show that stored mRNAs can be potentially more precise and quantitative biomarkers for monitoring seed aging of seed bank germplasm.

Results from this study show that stored mRNAs are degraded in the three plant species and their degradation in different plants share some similarities. This conclusion is supported by the observations: (1) △Ct value reflecting mRNA degradation showed a tight linear relationship with seed aging time for all three species, (2) different stored mRNAs analyzed for three species (six in Arabidopsis, four in wheat, and five in canola) showed constant degradation when seeds were aged under constant conditions, (3) RNA-seq analysis using MinION nanopore technique revealed a general decrease in almost all stored mRNA in aged Arabidopsis seeds. Results from the three different plants as well as recent results on soybean by others (Fleming *et al.,* 2017) suggest that stored mRNA degradation during seed aging is a general phenomenon for seeds of many plants.

Results from this study have revealed for the first time some fundamental features of stored mRNA degradation in plant seeds: (1) the similar β values over a wide range of aging time indicate a constant speed of degradation at the one nucleotide level when seeds are aged under constant conditions, (2) the "breaks" occur randomly along the entire length of a stored mRNA and (3) the β value or speed of stored mRNA degradation highly depends on the specific aging conditions such as temperature. The results also suggest that the stored mRNAs analyzed showed similar β values when seeds are aged under the same conditions. However, the number of stored mRNAs specifically analyzed only represents a very small portion of the total number of stored mRNAs present in dry seeds. Analysis of more stored mRNAs may reveal more clearly the extent of similarity and differences among different stored mRNAs in their degradation speeds.

One major conceptual advance made in this study is the creation of a new parameter, the β value, to quantitatively describe the average frequency of "breaks" per nucleotide per day in a stored mRNA under the constant aging conditions. The β value can allow quantitative comparisons of the rates of stored mRNA degradation among: (1) different genes of the same species, (2) different aging conditions, and (3) different species. Already, our results have revealed that the β value increased with aging temperature and accelerated after about 33 ^{o}C in Arabidopsis. Additionally, the β values for limited numbers of genes were quite close to each other among three species when seeds were aged under the similar conditions. It can be foreseen that this parameter will be useful for more comparative analyses in future studies.

5.1.3 Current findings challenge some traditional concepts on seed aging

The finding that seed stored mRNAs are degraded at relatively constant rates over the aging time for each of the three species in this study is quite intriguing and suggests a seed aging mechanism different from traditional ones. One of the classic theories in explaining seed aging is the gradual deformation of the glassy state which serves to stabilize molecule mobilization, limit damaging reactions, and prevent denaturation of macromolecules (Williams and Leopold, 1989). It is assumed that the mechanical properties of the glassy state succumb to seed storage (Bernal-Lugo and Leopold, 1998; Sun and Leopold, 1993). The glassy state is weakened because of the non-enzymatic deterioration of the macromolecules during seed aging (e.g. hydrolysis of oligosaccharides into reducing sugars) (Sun and Leopold, 1993, 1995). The weakened glassy state is less effective in limiting damaging reactions after storage and results in a sharp decrease in seed viability. Therefore, seed viability does not change much in the so-called asymptomatic phase, but then decreases sharply in the "rapid mortality" phase as observed in the present (Figure 3.11 for Arabidopsis, Figure 4.2 for wheat, and Figure 4.7 for *B.napus*) and other studies. Since the glassy state and molecular mobility should also affect mRNA degradation and if the glassy state in the dry seed is weakened gradually with the aging time, the rate of stored mRNA degradation should accelerate with the aging time as well. However, our results suggest that

stored mRNAs are degraded constantly over the aging time (as reflected by the β value) (Table 3.6), implying that the glassy matrix in limiting molecular mobility may not have changed significantly during seed aging under the present aging conditions, and the sigmoidal decrease of seed viability may have other causes.

Another theory is that seed deterioration is due to the accumulation of reactive molecules such as ROS (Bailly *et al*., 2008; Salvi *et al*., 2016a), reducing sugars (Walters, 2008), lipid peroxidation products (Marcos Filho, 2015; Seo *et al*., 2011), and denatured enzymes (Job *et al*., 2005). The ROS and other active molecules may cause seed deterioration independently or co-operatively (Kristal and Yu, 1992). Thus the accelerated loss of seed viability could be due to the increased levels of those reactive molecules as well as cooperation among them (Bernal-Lugo and Leopold, 1998). Since the rate of a chemical reaction depends on the concentrations of reactants, and if the levels of damaging reactive molecules such as ROS and reducing sugars are increasing during seed aging, the degradation reactions at the molecular level should show an increasing rate. The observation that stored mRNAs are degraded constantly is not consistent with this prediction of increased rates of degradation of macromolecules. It is possible that the increased ROS or other reactive molecules might be trapped inside the dry glassy matrix and only be set free during seed rehydration (Sun and Leopold, 1995). As a result, their increased levels in dry seeds during aging may not cause accelerated degradation of stored mRNAs. If so, their effects on other macromolecules may also be limited. Thus, it will be important to determine quantitatively the changes and particularly the rates of changes to other macromolecules during seed aging. With data from such analyses, we could then gain clearer and more mechanistic understanding regarding what happens at the molecular level and what the reason(s) may be for the sigmoidal loss of viability during seed aging.

5.1.4 Practical implications for germplasm management

The thorough characterization and better understanding of stored mRNA degradation in Arabidopsis, wheat, and canola allows the development of a more precise method (based on the $\triangle$Ct value of stored mRNAs) than the traditional methods. First, the $\triangle$Ct value for a

stored mRNA with a proper length is highly and linearly correlated with seed aging time linearly, and this quantitative relationship is valid for a wide time range. In contrast, other parameters either have a sigmoidal correlation to seed aging time (such as seed germination percentage and seedling weight) or have a much lower correlation coefficient (EC value). Further, a method based on the △Ct value can clearly detect the changes during the so-called asymptomatic phase of seed aging. Second, the △Ct value increases with fragment size, and thus different fragment lengths can be used to optimize sensitivity in practice. Such flexibility goes far beyond the capacity of traditional methods such as seed germination or the tetrazolium test. Third, the standard errors for the △Ct values in the current study are generally small even among biological repeats (Figure 3.9), indicating the high accuracy and repeatability of △Ct value. Thus, this method can provide a more precise and quantitative assessment of seed aging status than the traditional methods relying on seed germination which generally have high variance (Fu *et al*., 2017).

The methods we developed based on mRNA degradation could also facilitate germplasm management in several other aspects, such as assessing seed storability before storage and saving seed consumption. In germplasm conservation, seed storability is usually evaluated by an accelerated aging assay. The seeds that can withstand the AA treatment conditions longer are considered to have better storability – lasting longer during storage. The seed aging treatment and subsequent germination assay require at least several weeks to complete. Based on our study, mRNA degradation is relatively constant, and thus the level of mRNA integrity at the beginning of seed storage has the potential to be used as a good indicator of seed storability. In addition, its detection involves much less time. The second practical consideration is the seed consumption in the germination tests. To acquire reliable information on seed aging, hundreds of seeds are required for an accurate assessment on seed germination percentage. Consumption of hundreds of seeds for one test could be an important issue for seeds which are difficult to collect or have a big size, and it may be a more critical problem if multiple tests are needed before regeneration (Ayala *et al.*, 2000). In our analysis of wheat mRNA degradation, one single seed is enough for determining mRNA degradation although in the present study we used five seeds to minimize the variation among

different seeds. Clearly, a method based on stored mRNA degradation could save valuable seeds. To apply this method in monitoring seed aging, further optimization and standardization may be needed to suit the needs of specific plant species. Further, with the advancement of molecular techniques, more accurate or effective techniques e.g. droplet digital PCR (ddPCR) may be used.

5.2 Future work

5.2.1 Extending the studies to more plant species

In this study, it has been observed that stored mRNAs are degraded during seed aging in three plant species, suggesting that it may be a general phenomenon in orthodox seeds (i.e. seeds that can survive desiccation and/or freezing during *ex situ* conservation). Obviously, more plants need to be examined to confirm how universal this phenomenon is. The methods developed and findings made in this study will facilitate the studies in other plants. In addition, studies of some stored mRNAs in the three plants have revealed key fundamental features of stored mRNA degradation. First, the degradation or breakdown occurs randomly along the length of a stored mRNA. Second, the rate (or frequency) of breakdown for each nucleotide per day is constant under the same aging conditions. The new parameter, β value representing the probability of breakdown per nucleotide per day, should be very useful for comparing the rate of stored mRNA breakdown for different stored mRNAs (genes), under different aging conditions and among different plants. Such studies will provide a broader and deeper understanding regarding the extents of similarity and differences among different stored mRNAs and different plants.

The present results from three plant species show that the degradation of stored mRNAs as represented by ΔCt value is highly correlated with seed aging time. Thus, the level of stored mRNA (in the form of ΔCt value) can be used as a more precise marker of seed aging for many different plant species. Furthermore, in the present study, qPCR was mainly used to quantify the level of stored mRNAs. In further studies, more accurate techniques such as droplet digital PCR (ddPCR) can be tested and similarly used to quantify the stored mRNA levels and study seed aging.

5.2.2 Stored mRNAs and seed aging

It is generally believed that stored mRNAs are required at an early stage of seed germination since they provide the templates for synthesizing proteins before the *de novo* transcription is initiated in the seed. However, experimental evidence for this concept is based mostly on indirect results, for instance using inhibitors of protein synthesis and gene transcription (see Introduction and Literature Review). It will be interesting to understand the roles of certain families and those of specific stored mRNAs. The methods developed in this study would allow more precise and quantitative determination of specific stored mRNAs and could greatly facilitate the analyses on their potential roles in seed aging. On the other hand, certain Arabidopsis gene knockout and overexpressing mutants have been reported to have enhanced or reduced seed longevity (Seo *et al*., 2011; Mudgett *et al*., 1997; Oge *et al*., 2008; Chen *et al*., 2012). Analysis of stored mRNAs in these mutants will help to determine whether the degradation of stored mRNAs is correlated or involved in the seed longevity phenotypes of those mutants, and further to understand if and how any genetic factors could affect mRNA degradation during the seed aging process.

Reference

Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. *12*: R18.

Alkhalfioui, F., Renard, M., and Montrichard, F. (2007). Unique properties of NADP-thioredoxin reductase C in legumes. J. Exp. Bot. *58*, 969-978.

Allen, R.S., Li, J., Stahle, M.I., Dubroué, A., Gubler, F., and Millar, A.A. (2007). Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. Proc. Natl. Acad. Sci. U.S.A. *104*, 16371.

Almoguera, C., Prieto-Dapena, P., Diaz-Martin, J., Espinosa, J.M., Carranco, R., and Jordano, J. (2009). The HaDREB2 transcription factor enhances basal thermotolerance and longevity of seeds through functional interaction with HaHSFA9. BMC Plant Biol. *9*: 75.

Amuti, K.S., and Pollard, C.J. (1977). Soluble carbohydrates of dry and developing seeds. Phytochemistry *16*, 529-532.

Angelini, R., Cona, A., Federico, R., Fincato, P., Tavladoraki, P., and Tisi, A. (2010). Plant amine oxidases "on the move": an update. Plant Physiol. Biochem. *48*, 560-564.

Appelhagen, I., Lu, G.H., Huep, G., Schmelzer, E., Weisshaar, B., and Sagasser, M. (2011). TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of *Arabidopsis thaliana* seeds. Plant J. *67*, 406-419.

Aswad, D.W., Paranandi, M.V., and Schurter, B.T. (2000). Isoaspartate in peptides and proteins: formation, significance, and analysis. J. Pharm. Biomed. Anal. *21*, 1129-1136.

Ayala-Torres S., Chen, Y.M., Svoboda, T., Rosenblatt, J., and van Houten B. (2000). Analysis of gene-specific DNA damage and repair using quantitative polymerase chain reaction. Methods *22*, 135-147.

Bailly, C. (2004). Active oxygen species and antioxidants in seed biology. Seed Sci. Res. *14*, 93-107.

Bailly, C., El-Maarouf-Bouteau, H., and Corbineau, F. (2008). From intracellular signaling networks to cell death: the dual role of reactive oxygen species in seed physiology. C. R. Biol. *331*, 806-814.

Barciszewski, J., Barciszewska, M.Z., Siboska, G., Rattan, S.I.S., and Clark, B.F.C. (1999). Some unusual nucleic acid bases are products of hydroxyl radical oxidation of DNA and RNA. Mol. Biol. Rep. *26*, 231-238.

Basbouss-Serhal, I., Pateyron, S., Cochet, F., Leymarie, J., and Bailly, C. (2017). 5 ' to 3 ' mRNA decay contributes to the regulation of arabidopsis seed germination by dormancy. Plant Physiol. *173*, 1709-1723.

Bashkirov, V.I., Scherthan, H., Solinger, J.A., Buerstedde, J.M., and Heyer, W.D. (1997). A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates.

J. Cell Biol. *136*, 761-773.

Bazin, J., Langlade, N., Vincourt, P., Arribat, S., Balzergue, S., El-Maarouf-Bouteau, H., and Bailly, C. (2011). Targeted mRNA oxidation regulates sunflower seed dormancy alleviation during dry after-Ripening. Plant Cell *23*, 2196-2208.

Beelman, C.A., and Parker, R. (1995). Degradation of mRNA in eukaryotes. Cell *81*, 179-183.

Berjak, P., and Pammenter, N.W. (2008). From *Avicennia* to *Zizania*: seed recalcitrance in perspective. Ann. Bot. *101*, 213-228.

Berlett, B.S., and Stadtman, E.R. (1997). Protein oxidation in aging, disease, and oxidative stress. J. Biol. Chem. *272*, 20313-20316.

Bernal-Lugo, I., and Leopold, A.C. (1998). The dynamics of seed mortality. J. Exp. Bot. *49*, 1455-1461.

Bewley JD (1997) Seed germination and dormancy. Plant Cell *9*: 1055–1066

Boniecka, J., Kotowicz, K., Skrzypek, E., Dziurka, K., Rewers, M., Jedrzejczyk, I., Wilmowicz, E., Berdychowska, J., and Dabrowska, G.B. (2019). Potential biochemical, genetic and molecular markers of deterioration advancement in seeds of oilseed rape (*Brassica napus L.*). Ind. Crop. Prod. *130*, 478-490.

Bray, C.M., and Chow, T.Y. (1976). Lesions in the ribosomes of non-viable pea (*Pisum arvense*) embryonic axis tissue. Biochim. Biophys. Acta *442*, 14-23.

Britt, A.B. (1999). Molecular genetics of DNA repair in higher plants. Trends Plant Sci. *4*, 20-25.

Brocklehurst, P.A., and Fraser, R.S. (1980). Ribosomal RNA integrity and rate of seed germination. Planta *148*, 417-421.

Bruni, F., and Leopold, A.C. (1992). Pools of water in anhydrobiotic organisms - A thermally stimulated depolarization current study. Biophys. J. *63*, 663-672.

Bueso, E., Munoz-Bertomeu, J., Campos, F., Brunaud, V., Martinez, L., Sayas, E., Ballester, P., Yenush, L., and Serrano, R. (2014). *ARABIDOPSIS THALIANA HOMEOBOX25* uncovers a role for Gibberellins in seed longevity. Plant Physiol. *164*, 999-1010.

Buitink, J., and Leprince, O. (2004). Glass formation in plant anhydrobiotes: survival in the dry state. Cryobiology *48*, 215-228.

Buitink, J., and Leprince, O. (2008). Intracellular glasses and seed survival in the dry state. C. R. Biol. *331*, 788-795.

Cadet, J., Loft, S., Olinski, R., Evans, M.D., Bialkowski, K., Wagner, J.R., Dedon, P.C., Møller, P., Greenberg, M.M., and Cooke, M.S. (2012). Biologically relevant oxidants and terminology, classification and nomenclature of oxidatively generated damage to nucleobases and 2-deoxyribose in nucleic acids. Free Radical Res. *46*, 367-381.

Chatelain, E., Hundertmark, M., Leprince, O., Le Gall, S., Satour, P., Deligny-Penninck, S.,

Rogniaux, H., and Buitink, J. (2012). Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. Plant Cell Environ. *35*, 1440-1455.

Chatelain, E., Satour, P., Laugier, E., Vu, B.L., Payet, N., Rey, P., and Montrichard, F. (2013). Evidence for participation of the methionine sulfoxide reductase repair system in plant seed longevity. Proc. Natl. Acad. Sci. U.S.A. *110*, 3633-3638.

Chen, C.Y., and Shyu, A.B. (2011). Mechanisms of deadenylation-dependent decay. WIREs RNA *2*, 167-183.

Chen, H., Chu, P., Zhou, Y., Li, Y., Liu, J., Ding, Y., Tsang, E.W., Jiang, L., Wu, K., and Huang, S. (2012). Overexpression of AtOGG1, a DNA glycosylase/AP lyase, enhances seed longevity and abiotic stress tolerance in Arabidopsis. J. Exp. Bot. *63*, 4107-4121.

Chen, J.J., Chiang, Y.C., and Denis, C.L. (2002). CCR4, a 3 '-5 ' poly(A) RNA and ssDNA exonuclease, is the catalytic component of the cytoplasmic deadenylase. Embo J. *21*, 1414-1426.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: The exosome. Biochim. Biophys. Acta Gene Regul. Mech. *1829*, 552-560.

Close, T.J. (1996). Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. Physiol. Plantarum *97*, 795-803.

Comai, L., Dietrich, R.A., Maslyar, D.J., Baden, C.S., and Harada, J.J. (1989). Coordinate expression of transcriptionally regulated isocitrate lyase and malate synthase genes in *Brassica napus* L. Plant Cell *1*, 293-300.

Corbineau, F. (2012). Markers of seed quality: from present to future. Seed Sci. Res. *22*, S61-S68.

Cosio, C., and Dunand, C. (2009). Specific functions of individual class III peroxidase genes. J. Exp. Bot. *60*, 391-408.

Crowe, J.H., Crowe, L.M., Carpenter, J.F., and Wistrom, C.A. (1987). Stabilization of dry phospholipid-bilayers and proteins by sugars. Biochem. J. *242*, 1-10.

Crowe, L.M., and Crowe, J.H. (1992). Anhydrobiosis - a strategy for survival. Life sciences and space research Xxiv (3) : planetary biology and origins of life *12*, 239-247.

Dai, Y., Li, W., and An, L. (2016). NMD mechanism and the functions of Upf proteins in plant. Plant Cell Rep. *35*, 5-15.

David A.P. (1987). Seed aging. Implications for seed storage and persistence in the soil. Q. Rev. Biol. *62*, 196-197.

Davies, M.J. (2005). The oxidative environment and protein damage. Biochim. Biophys. Acta Proteins. Proteom. *1703*, 93-109.

de la Asuncion, J.G., Millan, A., Pla, R., Bruseghini, L., Esteras, A., Pallardo, F.V., Sastre, J.,

and Vina, J. (1996). Mitochondrial glutathione oxidation correlates with age-associated oxidative damage to mitochondrial DNA. FASEB J *10*, 333-338.

Debeaujon, I., Leon-Kloosterziel, K.M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. Plant Physiol. *122*, 403-413.

del Rio, L.A., Sandalio, L.M., Corpas, F.J., Palma, J.M., and Barroso, J.B. (2006). Reactive oxygen species and reactive nitrogen species in peroxisomes. Production, scavenging, and role in cell signaling. Plant Physiol. *141*, 330-335.

Devaiah, S.P., Pan, X.Q., Hong, Y.Y., Roth, M., Welti, R., and Wang, X.M. (2007). Enhancing seed quality and viability by suppressing phospholipase D in Arabidopsis. Plant J. *50*, 950-957.

Doetsch, P.W. (1995). What's old is new: an alternative DNA excision repair pathway. Trends Biochem. Sci. *20*, 384-386.

Drechsel, G., Kahles, A., Kesarwani, A.K., Stauffer, E., Behr, J., Drewe, P., Rätsch, G., and Wachter, A. (2013). Nonsense mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. Plant Cell *25*, 3726.

Dukan, S., Farewell, A., Ballesteros, M., Taddei, F., Radman, M., and Nystrom, T. (2000). Protein oxidation in response to increased transcriptional or translational errors. Proc. Natl. Acad. Sci. U.S.A. *97*, 5746-5749.

Dunckley, T., and Parker, R. (1999). The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. EMBO J. *18*, 5411-5422.

Dunlop, R.A., Rodgers, K.J., and Dean, R.T. (2002). Recent developments in the intracellular degradation of oxidized proteins. Free Radical Bio. Med. *33*, 894-906.

El-Maarouf-Bouteau, H., Meimoun, P., Job, C., Job, D., and Bailly, C. (2013). Role of protein and mRNA oxidation in seed dormancy and germination. Front. Plant Sci. *4*: 77.

Ellis, R.H., Hong, T.D., and Roberts, E.H. (1987). The development of desiccation-tolerance and maximum seed quality during seed maturation in six grain legumes. Ann. Bot. *59*, 23-29.

Ellis, R.H., and Roberts, E.H. (1981). The quantification of aging and survival in orthodox seeds. Seed Sci. Technol. *9*, 373-409.

Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A., and Mazzarella, L. (2000). The ultrahigh resolution crystal structure of ribonuclease A containing an isoaspartyl residue: Hydration and sterochemical analysis. J. Mol. Biol. *297*, 713-732.

FAO (2010). The second report on the state of the world's plant genetic resources for food and agriculture. Rome: FAO.

Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N., and Thurston, M. (2006). Open software for biologists: from famine to feast. Nat. Biotechnol. *24*, 801-803.

Finch-Savage W.E., Bassel G.W., Seed vigour and crop establishment: extending

performance beyond adaptation, J. Exp. Bot. 67, 567–591

Fleming, M.B., Hill, L.M., and Walters, C. (2018a). The kinetics of ageing in dry-stored seeds: a comparison of viability loss and RNA degradation in unique legacy seed collections. Ann. Bot. *123(7)*, 1133-1146.

Fleming, M.B., Patterson, E.L., Reeves, P.A., Richards, C.M., Gaines, T.A., and Walters, C. (2018b). Exploring the fate of mRNA in aging seeds: protection, destruction, or slow decay? J. Exp. Bot. *69*, 4309-4321.

Fleming, M.B., Richards, C.M., and Walters, C. (2017). Decline in RNA integrity of dry-stored soybean seeds correlates with loss of germination potential. J. Exp. Bot. *68*, 2219-2230.

Foyer, C.H., and Noctor, G. (2011). Ascorbate and glutathione: The heart of the redox hub. Plant Physiol. *155*, 2-18.

Frischmeyer, P.A., Van Hoof, A., Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science *295*, 2258.

Fu, Y.B., Ahmed, Z., and Diederichsen, A. (2015). Towards a better monitoring of seed ageing under *ex situ* seed conservation. Conserv. Physiol. *3*: cov026.

Fu, Y.B., Yang, M.H., Horbach, C., Kessler, D., Diederichsen, A., You, F.M., and Wang, H. (2017). Patterns of SSR variation in bread wheat (*Triticum aestivum* L.) seeds under *ex situ* genebank storage and accelerated ageing. Genet. Resour. Crop Evol. *64*, 277-290.

Fujiwara T, Eiji N., Yamagishi K, Goto D, Naito S. (2002) Storage proteins. In The Arabidopsis Book. *1*, e0020.

Gadaleta, M.N., Petruzzella, V., Renis, M., Fracasso, F., and Cantatore, P. (1990). Reduced transcription of mitochondrial DNA in the senescent rat. Eur. J. Biochem. *187*, 501-506.

Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D., and Rajjou, L. (2014). Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. Mol. Cell. Proteomics *13*, 252-268.

Galland, M., and Rajjou, L. (2015). Regulation of mRNA translation controls seed germination and is critical for seedling vigor. Front. Plant Sci. *6*: 284.

Ganguli, S., and Senmandi, S. (1993). Effects of aging on amylase activity and scutellar cell structure during imbibition in wheat seed. Ann. Bot. *71*, 411-416.

Garcia, D., Collier, S.A., Byrne, M.E., and Martienssen, R.A. (2006). Specification of leaf polarity in Arabidopsis via the trans-acting siRNA pathway. Curr. Biol. *16*, 933-938.

Garza-Caligaris, L.E., Avendano-Vazquez, A.O., Alvarado-Lopez, S., Zuniga-Sanchez, E., Orozco-Segovia, A., Perez-Ruiz, R.V., and Gamboa-Debuen, A. (2012). *At3g08030* transcript: a molecular marker of seed ageing. Ann. Bot. *110*, 1253-1260.

Gechev, T.S., Dinakar, C., Benina, M., Toneva, V., and Bartels, D. (2012). Molecular

mechanisms of desiccation tolerance in resurrection plants. Cell. Mol. Life. Sci. *69*, 3175-3186.

Geiger, T., and Clarke, S. (1987). Deamidation, isomerization, and racemization at Asparaginyl and Aspartyl residues in peptides - succinimide-linked reactions that contribute to protein-degradation. J. Biol. Chem. *262*, 785-794.

Goeres, D.C., Van Norman, J.M., Zhang, W., Fauver, N.A., Spencer, M.L., and Sieburth, L.E. (2007). Components of the Arabidopsis mRNA decapping complex are required for early seedling development. Plant Cell *19*, 1549-1564.

Golovina, E.A., Wolkers, W.F., and Hoekstra, F.A. (1997). Long-term stability of protein secondary structure in dry seeds. Comp. Biochem. Phys. A. *117*, 343-348.

Goyal, K., Walton, L.J., and Tunnacliffe, A. (2005). LEA proteins prevent protein aggregation due to water stress. Biochem. J. *388*, 151-157.

Gray, D. (1976). The effect of time to emergence on head weight and variation in head weight at maturity in lettuce (*Lactuca sativa*). Ann. App. Biol. *82*, 569-575.

Groot, S.P.C., de Groot, L., Kodde, J., and van Treuren, R. (2015). Prolonging the longevity of ex situ conserved seeds by storage under anoxia. Plant Genetic Resources-Characterization and Utilization *13*, 18-26.

Hamada, T., Tominaga, M., Fukaya, T., Nakamura, M., Nakano, A., Watanabe, Y., Hashimoto, T., and Baskin, T.I. (2012). RNA processing bodies, peroxisomes, golgi bodies, mitochondria, and endoplasmic reticulum tubule junctions frequently pause at cortical microtubules. Plant Cell Physiol. *53*, 699-708.

Hammond, S.M. (2005). Dicing and slicing. FEBS Lett. *579*, 5822-5829.

Hay FR, Probert RJ (2013) Advances in seed conservation of wild plant species: a review of recent research. Conserv Physiol. *1(1),* cot030.

Held, P. (2012). An introduction to reactive oxygen species. Tech Resources-App Guides *802*, 5-9.

Hoekstra, F.A., Golovina, E.A., and Buitink, J. (2001). Mechanisms of plant desiccation tolerance. Trends Plant Sci. *6*, 431-438.

Holdsworth, M.J., Finch-Savage, W.E., Grappin, P., and Job, D. (2008). Post-genomics dissection of seed dormancy and germination. Trends Plant Sci. *13*, 7-13.

Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. Cell *136*, 763-776.

Humphreys, A.M., Govaerts, R., Ficinski, S.Z., Lughadha, E.N., and Vorontsova, M.S. (2019). Global dataset shows geography and life form predict modern plant extinction and rediscovery. Nat. Ecol. Evol. *3*, 1043-1047.

Iwasaki, S., Takeda, A., Motose, H., and Watanabe, Y. (2007). Characterization of Arabidopsis decapping proteins AtDCP1 and AtDCP2, which are essential for post-

embryonic development. FEBS Lett. *581*, 2455-2459.

Ibl, V., and Stoger, E. (2012). The formation, function and fate of protein storage compartments in seeds. Protoplasma *249*, 379-392.

Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., Mueller, O., Schroeder, A., and Auffray, C. (2005). Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. Nucleic Acids Res. *33*, e56.

Januszyk, K., and Lima, C.D. (2014). The eukaryotic RNA exosome. Curr. Opin. Struct. Biol. *24*, 132-140.

Job, C., Rajjou, L., Lovigny, Y., Belghazi, M., and Job, D. (2005). Patterns of protein oxidation in Arabidopsis seeds and during germination. Plant Physiol. *138*, 790-802.

Johns, D.R. (1995). Mitochondrial DNA and disease. N. Engl. J. Med. *333*, 638-644.

Khah, E.M., Roberts, E.H., and Ellis, R.H. (1989). Effects of seed ageing on growth and yield of spring wheat at different plant-population densities. Field Crop. Res. *20*, 175-190.

Kimura, M., and Nambara, E. (2010). Stored and neosynthesized mRNA in Arabidopsis seeds: effects of cycloheximide and controlled deterioration treatment on the resumption of transcription during imbibition. Plant Mol. Biol. *73*, 119-129.

Koster, K.L. (1991). Glass-formation and desiccation tolerance in seeds. Plant Physio. *96*, 302-304.

Kranner, I., and Birtic, S. (2005). A modulating role for antioxidants in desiccation tolerance. Integr. Comp. Biol. *45*, 734-740.

Kranner, I., Birtić, S., Anderson, K.M., and Pritchard, H.W. (2006). Glutathione half-cell reduction potential: A universal stress marker and modulator of programmed cell death? Free Radical Bio. Med. *40*, 2155-2165.

Kranner, I., Chen, H.Y., Pritchard, H.W., Pearce, S.R., and Birtic, S. (2011). Inter-nucleosomal DNA fragmentation and loss of RNA integrity during seed ageing. Plant Growth Regul. *63*, 63-72.

Krishnan, P., Nagarajan, S., and Moharir, A. (2004). Thermodynamic characterisation of seed deterioration during storage under accelerated ageing conditions. Biosyst. Eng. *89*, 425-433.

Kristal, B.S., and Yu, B.P. (1992). An emerging hypothesis - synergistic induction of aging by free-radicals and Maillard reactions. J. Gerontol. *47*, B107-B114.

Kwon, Y.J., Park, M.J., Kim, S.G., Baldwin, I.T., and Park, C.M. (2014). Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. BMC Plant Biol. *14*, 136.

Łabno, A., Tomecki, R., and Dziembowski, A. (2016). Cytoplasmic RNA decay pathways - enzymes and mechanisms. BBA-Mol. Cell Res. *1863*, 3125-3147.

Lagrandeur, T.E., and Parker, R. (1998). Isolation and characterization of Dcp1p, the yeast

mRNA decapping enzyme. EMBO J. *17*, 1487-1496.

Lange, H., Zuber, H., Sement, F.M., Chicher, J., Kuhn, L., Hammann, P., Brunaud, V., Berard, C., Bouteiller, N., Balzergue, S.*, et al.* (2014). The RNA helicases AtMTR4 and HEN2 target specific subsets of nuclear transcripts for degradation by the nuclear exosome in *Arabidopsis thaliana*. PLoS Genet. *10(8)*. e1004564.

Lee, Y.P., Baek, K.H., Lee, H.S., Kwak, S.S., Bang, J.W., and Kwon, S.Y. (2010). Tobacco seeds simultaneously over-expressing Cu/Zn-superoxide dismutase and ascorbate peroxidase display enhanced seed longevity and germination rates under stress conditions. J. Exp. Bot. *61*, 2499-2506.

Leprince, O., Pellizzaro, A., Berriri, S., and Buitink, J. (2017). Late seed maturation: drying without dying. J. Exp. Bot. *68*, 827-841.

Leymarie, J., Vitkauskaite, G., Hoang, H.H., Gendreau, E., Chazoule, V., Meimoun, P., Corbineau, F., El-Maarouf-Bouteau, H., and Bailly, C. (2012). Role of reactive oxygen species in the regulation of Arabidopsis seed dormancy. Plant and Cell Physiol. *53*, 96-106.

Lezza, A.M., Boffoli, D., Scacco, S., Cantatore, P., and Gadaleta, M.N. (1994). Correlation between mitochondrial DNA 4977-bp deletion and respiratory chain enzyme activities in aging human skeletal muscles. Biochem. Biophys. Res. Commun. *205*, 772-779.

Liang, W., Li, C., Liu, F., Jiang, H., Li, S., Sun, J., Wu, X., and Li, C. (2009). The Arabidopsis homologs of CCR4-associated factor 1 show mRNA deadenylation activity and play a role in plant defence responses. Cell Res. *19*, 307-316.

Lima, J.J.P., Buitink, J., Lalanne, D., Rossi, R.F., Pelletier, S., da Silva, E.A.A., and Leprince, O. (2017). Molecular characterization of the acquisition of longevity during seed maturation in soybean. PLoS ONE *12(7):* e0180282.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods *25*, 402-408.

Loiseau, J., Vu, B.L., Macherel, M.H., and Le Deunff, Y. (2001). Seed lipoxygenases: occurrence and functions. Seed Sci. Res. *11*, 199-211.

Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinformatics *14*, 265-279.

Lykke-Andersen, J. (2002). Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. Mol. Cell. Biol. *22*, 8114-8121.

Maitra, R.D., Kim, J., and Dunbar, W.B. (2012). Recent advances in nanopore sequencing. Electrophoresis *33*, 3418-3428.

Mamula, M.J., Gee, R.J., Elliott, J.I., Sette, A., Southwood, S., Jones, P.J., and Blier, P.R. (1999). Isoaspartyl post-translational modification triggers autoimmune responses to self-proteins. J. Biol. Chem. *274*, 22321-22327.

Marcos F.J. (2015). Seed vigor testing: an overview of the past, present and future perspective. Sci. Agric. *72*, 363-374.

Marnett, L.J. (1999). Chemistry and biology of DNA damage by malondialdehyde. IARC Sci. Publ. *150:* 17-27.

Matthews, S., and Powell, A. (2006). Electrical conductivity vigour test: physiological basis and use. Seed Test. Int. *131*, 32-35.

Mayer, A.M., and Staples, R.C. (2002). Laccase: new functions for an old enzyme. Phytochemistry *60*, 551-565.

McDonald, M.B. (1999). Seed deterioration: physiology, repair and assessment. Seed Sci. Technol. *27*, 177-237.

Møller, I.M., Rogowska-Wrzesinska, A., and Rao, R.S.P. (2011). Protein carbonylation and metal-catalyzed protein oxidation in a cellular perspective. J. Proteomics *74*, 2228-2242.

Mornkham, T., Wangsomnuk, P.P., Fu, Y.B., Wangsomnuk, P., Jogloy, S., and Patanothai, A. (2013). Extractions of high quality RNA from the seeds of *Jerusalem Artichoke* and other plant species with high levels of starch and lipid. Plants (Basel) *2*, 302-316.

Moschou, P.N., Paschalidis, K.A., Delis, I.D., Andriopoulou, A.H., Lagiotis, G.D., Yakoumakis, D.I., and Roubelakis-Angelakis, K.A. (2008). Spermidine exodus and oxidation in the apoplast induced by abiotic stress is responsible for H_2O_2 signatures that direct tolerance responses in tobacco. Plant Cell *20*, 1708-1724.

Motomura, K., Le, Q.T., Kumakura, N., Fukaya, T., Takeda, A., and Watanabe, Y. (2012). The role of decapping proteins in the miRNA accumulation in Arabidopsis thaliana. RNA Biol. *9*, 644-652.

Motomura, K., Le, Q.T., Hamada, T., Kutsuna, N., Mano, S., Nishimura, M., and Watanabe, Y. (2014). Diffuse decapping enzyme DCP2 accumulates in DCP1 foci under heat stress in Arabidopsis thaliana. Plant Cell Physiol. *56*, 107-115.

Mudgett, M.B., Lowenson, J.D., and Clarke, S. (1997). Protein repair L-isoaspartyl methyltransferase in plants. Phylogenetic distribution and the accumulation of substrate proteins in aged barley seeds. Plant Physiol. *115*, 1481-1489.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and Bio assays with tobacco tissue cultures. Physiol. Plantarum *15*, 473-497.

Nagarajan, V.K., Jones, C.I., Newbury, S.F., and Green, P.J. (2013). XRN 5'→3' exoribonucleases: structure, mechanisms and functions. Biochim. Biophys. Acta *1829*, 590-603.

Nagel, M., Kranner, I., Neumann, K., Rolletschek, H., Seal, C.E., Colville, L., Fernandez-Marin, B., and Borner, A. (2015). Genome-wide association mapping and biochemical markers reveal that seed ageing and longevity are intricately affected by genetic background and developmental and environmental conditions in barley. Plant Cell Environ. *38*, 1011-1022.

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and

genetic regulation of transcription in seed. Plant J. *41*, 697-709.

Nash, H.M., Bruner, S.D., Scharer, O.D., Kawate, T., Addona, T.A., Spooner, E., Lane, W.S., and Verdine, G.L. (1996). Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. Curr. Biol. *6*, 968-980.

Negre-Salvayre, A., Coatrieux, C., Ingueneau, C., and Salvayre, R. (2008). Advanced lipid peroxidation end products in oxidative damage to proteins. Potential role in diseases and therapeutic prospects for the inhibitors. Br. J. Pharmacol. *153*, 6-20.

Nguyen, T.P., Keizer, P., van Eeuwijk, F., Smeekens, S., and Bentsink, L. (2012). Natural variation for seed longevity and seed dormancy are negatively correlated in Arabidopsis. Plant Physiol. *160*, 2083-2092.

Nishimura, N., Kitahata, N., Seki, M., Narusaka, Y., Narusaka, M., Kuromori, T., Asami, T., Shinozaki, K., and Hirayama, T. (2005). Analysis of ABA Hypersensitive Germination 2 revealed the pivotal functions of PARN in stress response in Arabidopsis. Plant J. *44*, 972-984.

Noctor, G., and Foyer, C.H. (1998). Ascorbate and glutathione: Keeping active oxygen under control. Annu. Rev. Plant Physiol. Plant Mol. Biol. *49*, 249-279.

Noctor, G., Mhamdi, A., and Foyer, C.H. (2016). Oxidative stress and antioxidative systems: recipes for successful data collection and interpretation. Plant Cell Environ. *39*, 1140-1160.

O'Brien, J.A., Daudi, A., Butt, V.S., and Bolwell, G.P. (2012). Reactive oxygen species and their role in plant defence and cell wall metabolism. Planta *236*, 765-779.

Oge, L., Bourdais, G., Bove, J., Collet, B., Godin, B., Granier, F., Boutin, J.P., Job, D., Jullien, M., and Grappin, P. (2008). Protein repair L-isoaspartyl methyltransferase 1 is involved in both seed longevity and germination vigor in *Arabidopsis*. Plant Cell *20*, 3022-3037.

Oikonomopoulos, S., Wang, Y.C., Djambazian, H., Badescu, D., and Ragoussis, J. (2016). Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. Sci. Rep. *6*: 31602.

Onate-Sanchez, L., and Vicente-Carbajosa, J. (2008). DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. BMC Res. Notes *1*: 93.

Oracz, K., Bouteau, H.E.M., Farrant, J.M., Cooper, K., Belghazi, M., Job, C., Job, D., Corbineau, F., and Bailly, C. (2007). ROS production and protein oxidation as a novel mechanism for seed dormancy alleviation. Plant J. *50*, 452-465.

Painter, R.B. (1985). Inhibition and recovery of DNA synthesis in human cells after exposure to ultraviolet light. Mutat. Res. *145*, 63-69.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature *425*, 257-263.

Pamplona, R. (2008). Membrane phospholipids, lipoxidative damage and molecular integrity: a causal role in aging and longevity. Biochim. Biophys. Acta *1777*, 1249-1262.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127-131.

Perry, D.A. (1984). Report of the vigor test committee 1980-1983. Seed Sci. Technol. *12*, 287-299.

Plooy, A.C.M., van Dijk, M., and Lohman, P.H.M. (1984). Induction and repair of DNA cross-links in Chinese hamster ovary cells treated with various platinum coordination compounds in relation to platinum binding to DNA, cytotoxicity, mutagenicity, and antitumor activity. Cancer Res. *44*, 2043-2051.

Pourcel, L., Routaboul, J.M., Kerhoas, L., Caboche, M., Lepiniec, L., and Debeaujon, I. (2005). *TRANSPARENT TESTA10* encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. Plant Cell *17*, 2966-2980.

Protic-Sabljic, M., and Kraemer, K.H. (1985). One pyrimidine dimer inactivates expression of a transfected gene in xeroderma pigmentosum cells. Proc. Natl. Acad. Sci. U.S.A. *82*, 6622-6626.

Pukacka, S., and Ratajczak, E. (2007). Ascorbate and glutathione metabolism during development and desiccation of orthodox and recalcitrant seeds of the genus *Acer*. Funct. Plant Biol. *34*, 601-613.

Rajjou, L., and Debeaujon, I. (2008). Seed longevity: survival and maintenance of high germination ability of dry seeds. C. R. Biol. *331*, 796-805.

Rajjou, L., Duval, M., Gallardo, K., Catusse, J., Bally, J., Job, C., and Job, D. (2012). Seed germination and vigor. Annu. Rev. Plant Biol. *63*, 507-533.

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The effect of α-Amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. Plant Physiol. *134*, 1598-1613.

Reiss, U., and Tappel, A.L. (1973). Fluorescent product formation and changes in structure of DNA reacted with peroxidizing arachidonic acid. Lipids *8*, 199-202.

Reuzeau, C., and Cavalie, G. (1997). Changes in RNA and protein metabolism associated with alterations in the germination efficiency of sunflower seeds. Ann. Bot. *80*, 131-137.

Rivett, A.J., and Levine, R.L. (1990). Metal-catalyzed oxidation of escherichia-coli glutamine-synthetase - correlation of structural and functional-changes. Arch. Biochem. Biophys. *278*, 26-34.

Roberts, B.E., Payne, P.I., and Osborne, D.J. (1973). Protein synthesis and the viability of rye grains. Loss of activity of protein-synthesizing systems *in vitro* associated with a loss of viability. Biochem. J. *131*, 275-286.

Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C., and Jaffe, D.B. (2013). Characterizing and measuring bias in sequence data. Genome Biol. *14*: R51.

Rushton, P.J., and Bray, C.M. (1987). Stored and *de novo* synthesized polyadenylated RNA

and loss of vigor and viability in wheat seed. Plant Sci. *51*, 51-59.

Salter, P.J., Currah, I.E., and Fellows, J.R. (2009). Studies on some sources of variation in carrot root weight. J. Agr. Sci. *96*, 549-556.

Salvi, P., Saxena, S.C., Petla, B.P., Kamble, N.U., Kaur, H., Verma, P., Rao, V., Ghosh, S., and Majee, M. (2016). Differentially expressed galactinol synthase(s) in chickpea are implicated in seed vigor and longevity by limiting the age induced ROS accumulation. Sci. Rep. *6*: 35088.

Sano, N., Ono, H., Murata, K., Yamada, T., Hirasawa, T., and Kanekatsu, M. (2015). Accumulation of long-lived mRNAs associated with germination in embryos during seed development of rice. J. Exp. Bot. *66*, 4035-4046.

Sano, N., Permana, H., Kumada, R., Shinozaki, Y., Tanabata, T., Yamada, T., Hirasawa, T., and Kanekatsu, M. (2012). Proteomic analysis of embryonic proteins synthesized from long-lived mRNAs during germination of rice seeds. Plant Cell Physiol. *53*, 687-698.

Sano, N., Rajjou, L., North, H.M., Debeaujon, I., Marion-Poll, A., and Seo, M. (2016). Staying alive: molecular aspects of seed longevity. Plant Cell Physiol. *57*, 660-674.

Sastre, J., Pallardo, F.V., and Vina, J. (2000). Mitochondrial oxidative stress plays a key role in aging and apoptosis. IUBMB Life *49*, 427-435.

Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M., and DellaPenna, D. (2004). Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. Plant Cell *16*, 1419-1432.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. *7*: 3.

Seo, Y.S., Kim, E.Y., and Kim, W.T. (2011). The *Arabidopsis sn*-1-specific mitochondrial acylhydrolase AtDLAH is positively correlated with seed viability. J. Exp. Bot. *62*, 5683-5698.

Shaban, M. (2013). Review on physiological aspects of seed deterioration. Intl. J. Agri. Crop Sci. *6*, 627-631.

Shan, X., Chang, Y.M., and Lin, C.L.G. (2007). Messenger RNA oxidation is an early event preceding cell death and causes reduced protein expression. Faseb J. *21*, 2753-2764.

Sharma, S.N., Maheshwari, A., Sharma, C., and Shukla, N. (2018). Gene expression patterns regulating the seed metabolism in relation to deterioration/ageing of primed mung bean (*Vigna radiata* L.) seeds. Plant Physiol. Biochem. *124*, 40-49.

Sheth, U., and Parker, R. (2006). Targeting of aberrant mRNAs to cytoplasmic processing bodies. Cell *125*, 1095-1109.

Shewfelt, R.L., and Purvis, A.C. (1995). Toward a comprehensive model for lipid-peroxidation in plant-tissue disorders. Hortscience *30*, 213-218.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nat. Struct. Mol. Biol. *19*, 594-601.

Spano, C., Bottega, S., Lorenzi, R., and Grilli, I. (2011). Ageing in embryos from wheat grains stored at different temperatures: oxidative stress and antioxidant response. Funct. Plant Bio. *38*, 624-631.

Sugliani, M., Rajjou, L., Clerkx, E.J., Koornneef, M., and Soppe, W.J. (2009). Natural modifiers of seed longevity in the Arabidopsis mutants *abscisic acid insensitive3-5* (*abi3-5*) and *leafy cotyledon1-3* (*lec1-3*). New Phytol. *184*, 898-908.

Sun, W.Q., and Leopold, A.C. (1993). The glassy state and accelerated aging of soybeans. Physiol. Plantarum *89*, 767-774.

Sun, W.Q., and Leopold, A.C. (1995). The Maillard reaction and oxidative stress during aging of soybean seeds. Physiol. Plantarum *94*, 94-104.

Sureshkumar, S., Dent, C., Seleznev, A., Tasset, C., and Balasubramanian, S. (2016). Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in Arabidopsis. Nat. Plants *2*, 16055.

Suzuki, Y., Ise, K., Li, C.Y., Honda, I., Iwai, Y., and Matsukura, U. (1999). Volatile components in stored rice [*Oryza sativa* (L.)] of varieties with and without lipoxygenase-3 in seeds. J. Agr. Food Chem. *47*, 1119-1124.

Suzuki, Y., Yasui, T., Matsukura, U., and Terao, J. (1996). Oxidative stability of bran lipids from rice variety [*Oryza sativa* (L.)] lacking lipoxygenase-3 in seeds. J. Agr. Food Chem. *44*, 3479-3483.

Tejedor-Cano, J., Prieto-Dapena, P., Almoguera, C., Carranco, R., Hiratsu, K., Ohme-Takagi, M., and Jordano, J. (2010). Loss of function of the HSFA9 seed longevity program. Plant Cell Environ. *33*, 1408-1417.

TeKrony, D.M., and Egli, D.B. (1991). Relationship of seed vigor to crop yield: A review. Crop Sci. *31*, 816-822.

Thompson, S., Bryant, J.A., and Brocklehurst, P.A. (1987). Changes in levels and integrity of ribosomal-RNA during seed maturation and germination in carrot (*Daucus Carota* L). J. Exp. Bot. *38*, 1343-1350.

Tuteja, N., Singh, M.B., Misra, M.K., Bhalla, P.L., and Tuteja, R. (2001). Molecular mechanisms of DNA damage and repair: progress in plants. Crit. Rev. Biochem. Mol. Biol. *36*, 337-397.

van Dijk, E.L., Chen, C.L., D'aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., Nicolas, A., Thermes, C., and Morillon, A. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114-117.

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. Trends Genet. *34*, 666-681

van Treuren R, de Groot EC, van Hintum JL (2013) Preservation of seed viability during 25 years of storage under standard genebank conditions. Genet. Resour. Crop Evol. *60*: 1407–1421.

van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262.

Vanessa O.M., Sidinei J.L., Solange B. T., Fernando A. H., Humberto D. Z., Liliane M. M. (2014). Cytogenetic analysis of wheat seeds submitted to artificial ageing stress. J. Seed Sci. *36*: 8.

Ventura, L., Dona, M., Macovei, A., Carbonera, D., Buttafava, A., Mondoni, A., Rossi, G., and Balestrazzi, A. (2012). Understanding the molecular pathways associated with seed vigor. Plant Physiol. Biochem. *60*, 196-206.

Walley, J.W., Kelley, D.R., Nestorova, G., Hirschberg, D.L., and Dehesh, K. (2010). Arabidopsis deadenylases AtCAF1a and AtCAF1b play overlapping and distinct roles in mediating environmental stress responses. Plant Physiol. *152*, 866.

Walters, C. (2008). Understanding the mechanisms and kinetics of seed aging. Seed Sci. Res. *8*, 223-244.

Walters, C. (2015). Genebanking seeds from natural populations. Nat. Area. J. *35*, 98-105.

Walters, C., Ballesteros, D., and Vertucci, V.A. (2010). Structural mechanics of seed deterioration: Standing the test of time. Plant Sci. *179*, 565-573.

Walters, C., Wheeler, L.M., and Grotenhuis, J.M. (2005). Longevity of seeds stored in a genebank: species characteristics. Seed Sci. Res. *15*, 1-20.

Waterworth, W.M., Masnavi, G., Bhardwaj, R.M., Jiang, Q., Bray, C.M., and West, C.E. (2010). A plant DNA ligase is an important determinant of seed longevity. Plant J. *63*, 848-860.

Weber, C., Nover, L., and Fauth, M. (2008). Plant stress granules and mRNA processing bodies are distinct from heat stress granules. Plant J. *56*, 517-530.

Weisshaar, B., and Jenkins, G.I. (1998). Phenylpropanoid biosynthesis and its regulation. Curr. Opin. Plant Biol. *1*, 251-257.

Williams, R.J., and Leopold, A.C. (1989). The glassy state in corn embryos. Plant Physiol. *89*, 977-981.

Wilson, D.O., and Mcdonald, M.B. (1986). The lipid-peroxidation model of seed aging. Seed Sci. Technol. *14*, 269-300.

Wolf, J., and Passmore, L.A. (2014). mRNA deadenylation by Pan2-Pan3. Biochem. Soc. Trans. *42*, 184-187.

Wolkers, W.F., Alberda, M., Koornneef, M., Leon-Kloosterziel, K.M., and Hoekstra, F.A. (1998). Properties of proteins and the glassy matrix in maturation-defective mutant seeds of *Arabidopsis thaliana*. Plant J. *16*, 133-143.

Wurr, J.R.F. (1983). The effect of the time of seedling emergence of crisp lettuce on the time of maturity and head weight at maturity. J. Hort. Sci. *58*, 561-566.

Xu, J., Yang, J.Y., Niu, Q.W., and Chua, N.H. (2006). *Arabidopsis* DCP2, DCP1, and VARICOSE form a decapping complex required for postembryonic development. Plant Cell *18*, 3386-3398.

Yakes, F.M., and VanHouten, B. (1997). Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. Proc. Natl. Acad. Sci. U.S.A. *94*, 514-519.

Yaklich, R.W., and Kulik, M.M. (1979). Evaluation of vigor tests in soybean seeds: relationship of the standard germination test, seedling vigor classification, seedling length, and tetrazolium staining to field performance. Crop Sci. *19*, 247-252.

Yang, H., Clendenin, W.M., Wong, D., Demple, B., Slupska, M.M., Chiang, J.H., and Miller, J.H. (2001). Enhanced activity of adenine-DNA glycosylase (Myh) by apurinic/apyrimidinic endonuclease (Ape1) in mammalian base excision repair of an A/GO mismatch. Nucleic Acids Res. *29*, 743-752.

Yoo, B.K., Santhekadur, P.K., Gredler, R., Chen, D., Emdad, L., Bhutia, S., Pannell, L., Fisher, P.B., and Sarkar, D. (2011). Increased RNA‐induced silencing complex (RISC) activity contributes to hepatocellular carcinoma. Hepatology *53*, 1538-1548.

Yoshida, R., Ogawa, Y., and Kasai, H. (2002). Urinary 8-oxo-7,8-dihydro-2'-deoxyguanosine values measured by an ELISA correlated well with measurements by high-performance liquid chromatography with electrochemical detection. Cancer Epidemiol. Biomarkers Prev. *11*, 1076-1081.

Zhang, X., and Guo, H. (2017). mRNA decay in plants: both quantity and quality matter. Curr. Opin. Plant Biol. *35*, 138-144.

Zhou, Y., Chu, P., Chen, H., Li, Y., Liu, J., Ding, Y., Tsang, E.W., Jiang, L., Wu, K., and Huang, S. (2012). Overexpression of *Nelumbo nucifera* metallothioneins 2a and 3 enhances seed germination vigor in Arabidopsis. Planta *235*, 523-537.

SUPPLEMENTARY FILE

Supplementary file 1: Perl scripts used in MinION nanopore sequencing

```
#!/usr/bin/perl
use warnings;
use strict;
my $file = '/home/liang/Desktop/usful_info_for_analysis/data_extracting/64allbwarna.sorted.sam';
open (ALIGNMENT_FILE, $file) or print "cannot open";
my @figure2_import_64 = <ALIGNMENT_FILE>;
close ALIGNMENT_FILE;
print "$figure2_import_64[$123456]";
my %gene_length;
my %gene_longest_read;
my %gene_read_count;
my $gene_count = 0;
my $readgene_count = 0;
foreach my $each_read_scalar (@figure2_import_64)   {
        my @each_read_array = split ('\s', $each_read_scalar);
            #HD
        if ($each_read_array[0]=~/^\@HD/)      {
                next;
                                                        }
            #SQ
        elsif ($each_read_array[0]=~/^\@SQ/)   {
            my $gene_name = substr($each_read_array[1],3,);
            print "$gene_name\n";
            my $gene_length_ = substr ($each_read_array[2],3,);
            $gene_length{$gene_name} = $gene_length_;
            print "$gene_length{$gene_name}\n";
            $gene_count++;
            next;
                                    }
              ###template reads
             elsif ($each_read_array[0] =~ /_Basecall_1D_template$/) {
                 my $key_name = $each_read_array[2];
                unless ($gene_length{$key_name}) {next;}
                if ($gene_longest_read{$key_name}) {$gene_read_count{$key_name}++; next;}
                else {
                      my $first_match_site = $each_read_array[3];
                    #  my $transcript_length = $gene_length{$key_name};
                      print "the first match point is $first_match_site and the gene length is
gene_length{$key_name} \n";
```

```
                my $longest_read_for_this_gene = $gene_length{$key_name} - $first_match_site ;
            $gene_longest_read{$key_name} = $longest_read_for_this_gene;
            print "$gene_longest_read{$key_name} \n";
            $readgene_count++;
            $gene_read_count{$key_name} = 1;
            }
                                                }
            }
print "total gene $gene_count total reads $readgene_count \n";

#####################################          AA part
my $file = '/home/liang/Desktop/usful_info_for_analysis/data_extracting/661+2allbwarna.sorted.sam';
open (ALIGNMENT_FILE, $file) or print "cannot open";
my @figure2_import_64 = <ALIGNMENT_FILE>;
close ALIGNMENT_FILE;
print "$figure2_import_64[$123456]";
my %gene_length_AA;
my %gene_longest_read_AA;
my %gene_read_count_AA;
my $gene_count_AA = 0;
my $readgene_count_AA = 0;
foreach my $each_read_scalar_AA (@figure2_import_64)   {
        my @each_read_array = split ('\s', $each_read_scalar);
            #HD
        if ($each_read_array[0]=~/^\@HD/)      {
                next;
                                                }
            #SQ
        elsif ($each_read_array[0]=~/^\@SQ/)   {
            my $gene_name = substr($each_read_array[1],3,);
            print "$gene_name\n";
            my $gene_length_ = substr ($each_read_array[2],3,);
            $gene_length{$gene_name} = $gene_length_;
            print "$gene_length{$gene_name}\n";
            $gene_count++;
            next;
                                    }
            elsif ($each_read_array[0] =~ /_Basecall_1D_template$/) {
                my $key_name = $each_read_array[2];
                unless ($gene_length{$key_name}) {next;}
                if ($gene_longest_read{$key_name}) {$gene_read_count{$key_name}++; next;}
                else {
                        my $first_match_site = $each_read_array[3];
```

```
                                print "the first match point is $first_match_site and gene length is $gene_length{$key_name} \n";
                                    my $longest_read_for_this_gene = $gene_length{$key_name} - $first_match_site ;
                            $gene_longest_read{$key_name} = $longest_read_for_this_gene;
                            print "$gene_longest_read{$key_name} \n";
                            $readgene_count++;
                            $gene_read_count{$key_name} = 1;
                            }
                                                        }
                            }
                               #  my $transcript_length = $gene_length{$key_name};
                                print "the first match point is $first_match_site and the gene length is gene_length{$key_name} \n";
                                my $longest_read_for_this_gene = $gene_length{$key_name} - $first_match_site ;

                            $gene_longest_read{$key_name} = $longest_read_for_this_gene;
                            print "$gene_longest_read{$key_name} \n";
                            $readgene_count++;
                            $gene_read_count{$key_name} = 1;
                            }
                                                        }
                            }
        print "total gene $gene_count total reads $readgene_count \n";

        ####################################          AA part
        my $file = '/home/liang/Desktop/usful_info_for_analysis/data_extracting/661+2allbwarna.sorted.sam';
        open (ALIGNMENT_FILE, $file) or print "cannot open";
        my @figure2_import_64 = <ALIGNMENT_FILE>;
        close ALIGNMENT_FILE;
        print "$figure2_import_64[$123456]";
        my %gene_length_AA;
        my %gene_longest_read_AA;
        my %gene_read_count_AA;
        my $gene_count_AA = 0;
        my $readgene_count_AA = 0;
        foreach my $each_read_scalar_AA (@figure2_import_64)   {
                my @each_read_array = split ('\s', $each_read_scalar);
                    #HD
                if ($each_read_array[0]=~/^\@HD/)      {
                        next;
                                                        }
                    #SQ
                elsif ($each_read_array[0]=~/^\@SQ/)   {
```

```
my $gene_name = substr($each_read_array[1],3,);
print "$gene_name\n";
my $gene_length_ = substr ($each_read_array[2],3,);
$gene_length{$gene_name} = $gene_length_;
print "$gene_length{$gene_name}\n";
$gene_count++;
next;
                                    }
elsif ($each_read_array[0] =~ /_Basecall_1D_template$/) {
    my $key_name = $each_read_array[2];
    unless ($gene_length{$key_name}) {next;}
    if ($gene_longest_read{$key_name}) {$gene_read_count{$key_name}++; next;}
    else {
        my $first_match_site = $each_read_array[3];
                print "the first match point is $first_match_site and gene length is
$gene_length{$key_name} \n";
        my $longest_read_for_this_gene = $gene_length{$key_name} - $first_match_site ;
    $gene_longest_read{$key_name} = $longest_read_for_this_gene;
    print "$gene_longest_read{$key_name} \n";
    $readgene_count++;
    $gene_read_count{$key_name} = 1;
    }
                                        }
    }
```